Evaluating Cost Effectiveness of the USDA's Expanded Food and Nutrition Education Program

Ranju Baral

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Economics, Agriculture

George C. Davis, Chair

Elena Serrano

Wen You

Gary Skaggs

Bradford Mills

June 6, 2013

Blacksburg, Virginia

Keywords: Expanded Food and Nutrition Education Program, cost effectiveness

Evaluating Cost Effectiveness of the USDA's Expanded Food and Nutrition Education Program

Ranju Baral

ABSTRACT

The Expanded Food and Nutrition Education Program (EFNEP) is one of the largest efforts of the US Department of Agriculture (USDA) to promote healthy dietary behavior practices among the low income adults and youths in the US. Although the program is shown to be effective in achieving its stated goals, the cost effectiveness of the money spent on EFNEP remains largely unknown. This dissertation analyzes the costs and effectiveness of the EFNEP, and is organized in three essays. The first essay investigates the effectiveness of the adult EFNEP and evaluates the returns to scale on the money spent in this program by utilizing an indirect production function approach. Results indicate that the program has increasing returns to scale at the National level, although a significant variation exists across the states. The second essay develops a framework for conducting the cost effectiveness analysis (CEA) for the youth EFNEP. The CEA framework is then applied to the data from Virginia youth EFNEP to estimate the cost effectiveness ratio (CER). The CER is estimated to be about $75 per behavior improvement. The third essay examines the outcomes and the attributes of the youth EFNEP program using the Rasch model type measurement model. Findings suggest that the youth EFNEP is effective in achieving its stated program goals. In addition, the program related characteristics are found to be important attributes of effective programs. Overall, this dissertation has important policy implications for improving the (cost) effectiveness of nutrition education programs.

DEDICATION

To my daughters, *Aumi* and *Kaya,* who sacrificed their cherished childhood to let me accomplish my academic goals.

ACKNOWLEDGEMENTS

I am indebted to my advisor Dr. George C. Davis. I would not have completed this task without his incessant support and encouragement. He has helped me grow as a scholar and inculcated in me the importance of scientific inquiry and publication. I am very thankful to his time and enthusiasm to meticulously explain to me some of the basic concepts in economics. His focus on simplicity has changed the way I view the real world problems. My academic journey at VT has been very rewarding because of him. I have no words to thank you Dr. Davis.

I would like to thank all my dissertation committee members Dr. Elena Serrano, Dr. Wen You, Dr. Gary Skaggs, and Dr. Bradford Mills for their support and inputs at various stages of my dissertation. Dr. Elena Serrano was instrumental in developing and executing the cost effectiveness analysis project. Without her help, the project would not have taken off. Dr. Wen You provided valuable inputs for the econometric analysis. I hope I didn't bore you with my silly questions. Whenever I needed help in the measurement model, Dr. Gary Skaggs provided valuable guidance. I greatly appreciate his inputs during the development and validation phase of the survey instruments. Dr. Bradford Mills offered important insights to improve the quality of research during the committee meetings.

I am very thankful to Ms. Stepahnie Blake, Program Specialist for the EFNEP at USDA, for providing me the data on adult EFNEP program. She has always been very prompt in responding my queries and has helped me understand the NEERS dataset. It was a great experience collaborating with you. Virginia EFNEP team deserves special thanks. Dr. Mary McFerren has been very kind in extending her help and has facilitated the whole process of data collection by bridging us to other EFNEP staffs and family. EFNEP staffs, area coordinators, program assistants, unit coordinators and all staffs at the county level office have helped us in

data collection. I cannot overstate the hard work of the program assistants who have been directly involved in data collection and for sending the data on time. Without your collaboration, this dissertation would have been incomplete. I also thank all the participants of youth EFNEP program for taking time to fill out surveys.

My friends deserve special thanks for making my stay at VT enjoyable. I could not be more grateful to Niraj Poudyal for all his help to teach me complex economic equations. Never before meeting Niraj had I experienced that learning from peers could have substantial impacts for scholarship. I am indebted to Sapna Kaul and Kristen Capogrossi for being true friends with whom I shared my joys and frustrations. With my colleagues Amal Almahanna, and Keeletlhoko Faith Badirwang, I have shared the joy of entering the data.

I have to admit that my PhD journey began without much forethought. The courage my husband Nabin provided is beyond my expectations. He has taught me to smile at hard times. He has been my mentor who I could count on; he has been my friend who I could rely on. My life has been much better because of you. I admire having you in my life more than you know.

Last but not the least I am thankful to have such a wonderful family, my parents, my sisters and brother and my in-laws, who supported me unconditionally through ups and downs.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

Chapter 1: Introduction

1.1. Background and Motivation

Nutrition education programs are a common policy tool for improving nutrition behaviors and consequently public health. Nutrition education has become more critical in the context of an obesity epidemic in the United States. The increasing trend of overweight and obese people shows a need for effective and far-reaching nutrition education programs targeted towards the current and future populations at risk. The US government's commitment for improving the nutritional health of its citizens is reflected in the federal funding of a breadth of programs devoted to that cause. The US Department of Agriculture (USDA) spends millions of dollars every year [$472 million in fiscal year 2002 (GAO 2004)], on nutrition education programs such as the Expanded Food and Nutrition Education Program (EFNEP); Women, Infants and Children (WIC) Program; and Supplemental Nutrition Assistance Program (SNAP-ED), to help ensure the nutritional wellbeing of thousands of people nationwide. This study focuses on the EFNEP.

The EFNEP is designed specifically to provide nutrition education, whereas the others combine nutrition education together with nutrition assistance programs (GAO 2004). The goal of EFNEP is "to assist low-income families and youths in acquiring the knowledge, skills, attitudes, and changed behavior necessary for nutritionally sound diets, and to contribute to their personal development and the improvement of the total family diet and nutritional well-being" (USDA EFNEP Policy Paper 1983). With its history of more than 40 years, it now operates in more than 800 counties of all the 50 states and 6 US territories (USDA 2011a). Originally conceived to deal with the problem of undernourishment, the nutrition education programs have morphed to abate the problems of overweight and obesity lately. Lower socioeconomic status

populations are usually targeted because of the well-established positive relationship between the socioeconomic status and health quality (e.g., Marmot and Wadsworth 1999).

The Expanded Food and Nutrition Education Program (EFNEP), one of the largest federally funded nutrition education programs in the United States spends about $68 million of federal budget every year (NIFA 2011). Every year, EFNEP serves more than a half million people. In the Fiscal Year 2010/11 alone, it served about 506,156 youths and 134,446 adults. Youth EFNEP is designed for youths aged 2 to 19 years and is delivered in different forms such as at schools as an enrichment of the curriculum, in after-school care programs through 4-H EFNEP clubs, camps, home gardening workshop (USDA 2011a). Youth programs focus on lessons on nutrition, food preparation, and food safety, and related topics, including physical activity and health. Adult EFNEP serves adults aged more than 19 years and is delivered in small groups or at an individual level. Adult programs focus on improving various aspects of nutrition ranging from participants' ability to make healthy food choices, skills in food preparation, safety and sanitation to managing food budget. Trained paraprofessionals called program assistants (PAs), and volunteers deliver both youth and adult EFNEP programs as a series of lessons over several months.

Despite the long history of nutrition education programs, very little is known about the costs and effectiveness of such programs in ensuring the nutritional health of people they serve. The General Accounting Office 2004 report for the accountability and efficiency of federally funded programs recognized that there is a lack of economic evaluation tools being used in the federally-funded nutrition education programs and made recommendations for pursuing such endeavor as a part of ensuring accountability of programs (GAO 2004). In today's tight budgetary environment, evidence of program's cost effectiveness in achieving its stated goals

could provide justification for its appropriation. This dissertation is an attempt to conduct an economic evaluation of the EFNEP with the aim of gleaning potential evidence for cost effectiveness of the EFNEP.

EFNEP Effectiveness and Gaps in the Literature

As a routine program evaluation, the USDA publishes the annual impact reports of EFNEP. For the evaluation of adult EFNEP, USDA uses responses to a 10 question "behavior checklist" to construct three "impact" indicators: (i) food resource management practices (FRMP), (ii) nutrition practices (NP) and (iii) food safety practices (FSP). They aggregate the number of adult participants who showed improvement in one or more behavior indicators from a *pre-test* to a *post-test* and use it as a measure of adult EFNEP effectiveness. The evaluation of youth EFNEP is primarily based on the following four impact indicators: (i) increase in variety of foods they eat; (ii) increase in nutritional knowledge; (iii) increase in ability to select low-cost, nutritious food; and (iv) improvement in food preparation and safety practices. The numbers of youths who improve in each of these indicators are reported as impacts of youth program by the USDA.

USDA's annual impact report shows that most EFNEP participants (both adults and youths) tend to improve in different domains of nutrition behavior (USDA 2011a). The program is effective in achieving its goal of improving nutrition behavior among its participants (USDA 2011a). Evaluation of the EFNEP effectiveness by USDA is largely based on simple aggregates of behavior indicators reported in the annual report. The impact data, however, are only

summary statistics on the outcomes with no multivariate analysis associating these results with dollar expenditures, and program, or participant characteristics.

To address the well-known limitations of the aggregate summary statistics available from the USDA, several studies have documented the positive behavior changes and cost effectiveness of adult EFNEP in specific states. The cost benefit analysis of Virginia EFNEP for the year 1996 was apparently the first such study to be published in a peer reviewed journal (Rajgopal et al. 2002). They found the benefit to cost ratio for the adult EFNEP to be $10.64/$1.00 (Rajgopal et al. 2002). A couple of other states followed the suit to estimate the benefit to cost ratio for their respective programs. The estimates of the benefit to cost ratio was $8.34 for California (Joy and Goldman 2006), $12.50 for Iowa (Wessman and Jensen 2002), $3.62 for Oregon (Schuster et al. 2003), and $9.58 for New York (Dollahite et al. 2008). Variations in the estimates of benefit to cost ratios across states are due to the variation in assumption regarding the measure of costs and benefits across these studies. The New York study also conducted a cost effectiveness analysis in addition to the cost benefit analysis. For New York EFNEP, the program cost was US $892 per graduate and the cost effectiveness ratio was $20,863 per quality adjusted life year (QALY) saved (Dollahite et al. 2008).

While the existing results on the effectiveness of adult EFNEP are encouraging and suggestive, generalizability of these studies is restricted due to their limited scope in sample size, studying a single state, and using data for a single year. There are still several outstanding questions related to the effectiveness of the money spent on the adult EFNEP. Whether the money spent on the adult EFNEP actually contributes to the stated objectives of the program is unknown at the national level. Further, the relationship between the amount of money spent on the program and its impact remains unexplored. There is no indication of whether the variation in

the allocation of money across programs has any impact on the adult EFNEP program's effectiveness.

Compared to the adult EFNEP, studies on youth EFNEP are even sparser, although youths consist of 75% of the EFNEP beneficiaries (USDA 2011a). Townsend and colleagues (2006) are only scholars to study the effectiveness of youth EFNEP. By using USDA's selected youth impact indicators as outcome measures, they found that the youth participants who received the lessons had higher post-test scores than their counterparts. Townsend et al. (2006), however, did not look at the costs of the program.

Part of the reason for the scantiness in evaluation studies on youth EFNEP is due to the lack of valid and reliable assessment tools. No federal guideline mandates the use of any specific instrument or even curricula for youth EFNEP. For evaluation, the state programs often create their own survey instruments. This results in the lack of consistency and standardization across youth EFNEP program evaluation.

Given there is really only one study of the effectiveness of youth program (Townsend et al. 2006) it is perhaps no surprise that there has been no cost effectiveness analysis of youth programs. There are no studies that analyze the cost and benefit of youth EFNEP. The cost benefit studies similar to the ones done for adults are more challenging for the youths, especially because of the multitude of factors that can influence long-term health and disease outcomes. Measuring the total monetary value of the program benefits accrued by youths could be very imprecise. Cost effectiveness analysis could be an alternative to cost benefit analysis, which does not require converting the benefits into monetary units (Gold et al. 1996), but is equally informative in stating the program success in achieving its goal.

As discussed above, some of the basic questions related to the cost and effectiveness of the EFNEP have not received much attention in the literature, and it provides motivation for this dissertation. The state of knowledge about the effectiveness and efficiency of adult and youth program is different. For adult EFNEP there are a couple of studies which investigate the cost effectiveness ratios of adult programs whereas such studies are nonexistent for youth programs. The relevant question for the adult EFNEP, then, is to ask how effective the program is at the national level, how performances of different states compare at the National level. Such quest is a step ahead for the youth EFNEP. A more relevant question for the youth program would be to ask how much it costs the youth program to achieve its stated objectives, in other words the cost effectiveness ratios. This dissertation addresses these questions related to adult and youth programs, separately.

1.2. Research Goals

The overall goal of this dissertation is to investigate the cost effectiveness of the EFNEP in reaching the program objectives and to explore the factors that contribute to the effectiveness of the program. The specific research goals of this dissertation are as follows

Goal 1: To explore the effectiveness of money spent on adult EFNEP at the National Level

Goal 2: To conduct a cost effectiveness analysis of the youth EFNEP in Virginia

Goal 3: To analyze the outcomes of the youth EFNEP program in Virginia and explore the factors that contributes to its effectiveness

Goal 1 is concerned with the evaluation of the adult EFNEP in achieving its goals at the national level. To accomplish this goal, the Nutrition Education Evaluation and Reporting System (NEERS) data on program outcomes and other covariates, and the annual allocation data for all 50 states and 6 US territories for the years 2000–2006 are used. Using an indirect production function approach, the returns to scale on the money spent on the program with respect to three outcome indicators, as used by the USDA, are evaluated for all states and at the national level. We find that the adult program has increasing and constant returns to scale for two of the three outcome indicators, at the National level, although a significant variation exists across the states. Further, program and participant characteristics do not seem to be as important as the amount of money spent on the program in determining the program outcomes.

Goals 2 and 3 are concerned with the evaluation of the youth EFNEP program at the state level. Interest in goal 2 is to develop a framework for conducting a cost effectiveness analysis (CEA) for the youth program. With a concerted effort of the advisory board members (which consist of faculties and experts in EFNEP, program evaluation, nutrition, and economics) a framework for conducting a CEA is developed. A comprehensive and validated tool for measuring the effects of youth EFNEP is developed. The framework is then applied to get the estimates of the cost effectiveness ratios of youth EFNEP for Virginia. Primary data on program outcomes are collected from program participants and the costs data are collected from program managers from 15 counties in Virginia who implemented the youth program in the year 2011/12, and used in the analysis. The estimate of CER is found to be around $75 per improvement in behavior outcome. A wide variation in the estimates of CER across counties is found.

Goal 3 specifically is concerned with analyzing the outcomes of the youth EFNEP in Virginia and in identifying the attributes of program effectiveness. To achieve this objective,

primary data collected for goal 2 is used. A different perspective to measure program outcomes is followed, where the outcomes are viewed as latent variables unlike a traditional approach used by the USDA (which is to count the number of individuals who improve in at least one question). The Rasch model type measurement model is used to accomplish goal 3. We find that overall the youth EFNEP in Virginia is effective in improving the nutrition behavior among its participants. Several of the program and participants characteristics such as participant's age, class size, program duration, program assistant's experience, are identified as important attributes of effective program.

Answers to the basic questions related to costs and effectiveness of EFNEP is of significant interest to EFNEP policy makers at the National level for directing inquiries into which states are more efficient. Similarly, the procedures demonstrated for conducting the cost effectiveness analysis are useful and appealing for the EFNEP administrators at the State level as they are practical and simple. This can potentially guide the EFNEP administrators to identify programs that are doing relatively better than the others. Identification of attributes of effective program can provide guidance to devise effective program policies, which can ultimately contribute towards improving the overall EFNEP effectiveness.

An implicit assumption throughout this dissertation is that the program effectiveness is accurately captured by the survey data used in the analysis. The survey data used here is consistent with that used in the nutrition literature.

1.3. Organization of Dissertation

This dissertation is organized in five chapters. Each of the three goals is organized as three separate essays. Chapter 1 (this chapter) is the introduction chapter and presents the background and motivation for evaluating the USDA's EFNEP program. The gaps in the literature and how this dissertation contributes to this gap are explained. Chapter 2 includes an essay related to the evaluation of the adult EFNEP at the National level. This paper has already been published in the *Journal of Agricultural and Applied Economics*, volume 45, issue 2, year 2013, and is included in the exact form as it is published in the journal. Chapter 3 is an essay on the estimates of the cost effectiveness ratios of the youth EFNEP program in Virginia. Chapter 4 is an essay that investigates outcomes of the youth EFNEP and explores the factors affecting the program effectiveness. Both chapters 3 and 4 are being developed as separate manuscripts and will be submitted for publication in the peer reviewed journal. Finally, chapter 5 is a conclusion chapter which summarizes major research findings, research implications, some of the limitations, and potential avenues for future research.

Chapter 2: National, Regional, and State Level Estimates of Returns to Scale in the Expanded Food and Nutrition Education Program

Abstract. The effectiveness of the Expanded Food and Nutrition Education Program (EFNEP) in achieving its goals at the national, regional, and state level is unknown. Utilizing USDA data from all states and territories for the years 2000-2006, the impact of program and participant characteristics and returns to scale on the three outcome indicators used by the USDA are estimated. Program and participant characteristics do not seem to be as important as the amount of money spent on the program. Generally speaking, there are constant and increasing returns to scale for two of the three federal outcome indices for most states, but not all.

Baral, R., G.C. Davis, and W. You (2013). National, Regional, and State Level Estimates of Returns to Scale in the Expanded Food and Nutrition Education Program. Journal of Agricultural and Applied Economics, 45(2):203-216

2.1. Background

Nutrition education programs are a common policy tool for improving nutrition and consequently public health. Lower socioeconomic status (SES) populations are usually targeted because of the well-established positive relationship between SES and health quality (e.g., Marmot and Wadsworth 1999). The Expanded Food and Nutrition Education Program (EFNEP) is one of the largest federally funded nutrition education programs in the United States and is administered by the United States Department of Agriculture (USDA) (General Accounting Office 2004). In operation for more than 40 years and now in all 50 states and 6 territories, the EFNEP has become a cornerstone in US nutrition education (USDA 2009a).

The aim of the EFNEP is to assist limited-resource audiences to "acquire the knowledge, skills, attitudes, and changed behavior necessary for nutritionally sound diets, and to contribute to their personal development and improvement of the overall family diet and nutritional well-being" (USDA 2009a). The EFNEP is administered by the USDA Cooperative Extension Service to youths and adults. The Adult EFNEP serves adults 19 years and older and is the focus in this paper.

Within each state, the State Cooperative Extension staffs provide leadership in the program and choose and/or develop the curriculum. The curriculum primarily focuses on improving various aspects of nutrition ranging from participants' ability to make healthy food choices, skills in food preparation, food safety and sanitation to managing a food budget. The State Cooperative Extension staffs provide training to the paraprofessionals (also called program assistants, PAs), and volunteers, who are usually local to the community they serve, and are responsible for recruiting the participants and delivering the EFNEP curriculum. The curriculum is delivered as a series of lessons, at least six as required by the USDA, over several months. The

number of lessons and the topics covered in the lessons are usually tailored to the needs of the program participants and usually vary across the states. The program is delivered to individuals or to a group of individuals depending on the need and convenience.

Though USDA spends about $66 million per year on the EFNEP (USDA 2009b), very little is known about the impact of this money on achieving its stated objectives at the *national* level. There are a few single-state, single-year analyses for subsets of EFNEP participants (e.g., Arnold and Sobal 2000; Dickin, Dollahite and Habicht 2005; Dollahite and Pierce 2003) and the general finding is that the EFNEP improves nutrition knowledge and behavior, but these analyses cannot be generalized to a national level. At the national level the only information provided is by the USDA's EFNEP "impact data" annual report, where impact is measured as the percentage of participants showing improvement from a *pre-test* to a *post-test* on questions related to food resource management practices, nutrition practices, and food safety practices (USDA 2009a). The impact data suggest most participants tend to improve in these domains. However, the impact data are only summary statistics on the outcomes with no multivariate analysis associating these results with dollar expenditures and program/participant characteristics.

While the few state level results are encouraging and suggestive, there are still several outstanding questions related to the effectiveness of the money spent on the EFNEP program at the national level:

(1) Are there program or participant characteristics that can be identified as contributing to the stated objectives?

(2) Does the money spent on the EFNEP contribute to the stated objectives of the program?

(3) What are the returns to scale associated with the EFNEP money in achieving its objectives at the national level as well as at the state/territory level?

(4) Are there geographical differences across states and regions in the returns to scale measures?

Using the USDA impact data for all states and US territories for seven consecutive years (2000–2006), but in conjunction with several available covariates, we answer these questions using a nonlinear multivariate regression model. Results show that only a few program and participant characteristics are significant in explaining changes in the outcome indices used by USDA. The money spent on the EFNEP is the most important factor and has a positive and significant impact on two of the three outcome indices. Using a purely statistical significance criterion, there is constant returns to scale at the national level for two of the three indices (food resource management practices and nutrition practices), whereas three of the five regions show increasing returns in the food resource management index. About 40 percent of states/territories show increasing returns to scale in one of the two indices, and the remaining states mainly show constant returns to scale in both the indices. The third index (i.e., food safety practices) shows decreasing returns to scale regardless of the level of analysis. A broader perspective based on economic significance versus statistical significance suggests that a majority of the states have increasing returns to scale in two of the three indices. These findings are the first to quantify the budget effects at the national level and also help in identifying potential sources where cost effective information may be gleaned. Limitations are discussed in the conclusions.

2.2. Conceptual Framework

When evaluating publicly funded projects, Shephard (1974) advocates the use of an indirect production function approach. The indirect production function comes from maximizing output subject to a cost constraint as opposed to the common dual problem of minimizing expenditures subject to a production constraint that generates the cost function. In a public funding setting, such as the EFNEP, the constraint is the amount of money provided to operate the program and given that amount of money, the goal is to maximize the output, which in this case would be positive outcome measures of the program. Thus the indirect production function can be written as $Y = f(B, \boldsymbol{Z})$, where Y is a measure of output, B is the budget amount and $\boldsymbol{Z}$ is a vector of other covariates that may affect the outcome such as participant and program characteristics. Given the functional form that will be used, the marginal effect of the budget on the output turns out to be a measure of returns to scale, as will be shown.

2.3. Data

USDA collects data on the EFNEP by state/territory through its Nutrition Education Evaluation and Reporting System (NEERS) (USDA 2009c). The NEERS is a software program that allows administrators at the county and state level to collect and report data to USDA in aggregate form on adult participation. Given the structure of the program, the budget effects potentially interact with participant and program characteristics so the analysis must incorporate these potential confounding factors. In general, the NEERS collects data on participant characteristics (e.g., age, education, and income), program characteristics (e.g., type of lesson, type of instructor) and most importantly responses to 10 required behavioral checklist questions

for adults. The total EFNEP budget for each state/territory is available as well. The data covers all 50 states and 6 US territories for the fiscal years 2000 to 2006.

Dependent Variables

We use the same measurements developed by USDA as our dependent variables. Specifically, the USDA uses 10 behavioral checklist questions to form the basis for evaluating the success of the EFNEP. The 10 behavioral checklist questions were developed over a span of five years from 1993 to 1997 and involved several phases of development that assured content and face validity (see Anliker, Willis, and Montgomery 1998 for details). The major phases were (i) a committee formed the questions based on reviews of other existing and submitted instruments, (ii) feedback was solicited on the instrument from the EFNEP Coordinators in all 50 states, (iii) more feedback on the instrument was solicited from a larger pool of specialists, (iv) an Expert Panel met in Washington DC to revise the instrument, (v) the revised instrument was subjected to focus group testing and tested for reading level (determined it was at the 6^{th} grade reading level), (vi) based on feedback from the focus group, the instrument was revised and pilot tested in seven states, (vii) statistical analysis was conducted on the pilot tested instrument and the instrument was checked for validity and internal reliability, and (viii) additional minor revisions were made based on the validity and reliability analysis to improve and finalize the instrument.

The 10 behavioral questions are given in table 1 and responses are based on a 5 point Likert-type scale (e.g., 1 = do not do, 2 = seldom, 3 = sometimes, 4 = most of the times, 5 = almost always). Participants answer these behavioral questions both pre and post EFNEP participation. USDA uses subsets of the 10 questions to create three composite measures of

different aspects of nutrition knowledge: the Food Resource Management Practices (FRMP) composite is constructed from answers to questions 1–4, the Nutrition Practices (NP) composite is constructed from answers to questions 1 and 7–10, and the Food Safety Practices (FSP) composite is constructed from answers to questions 5 and 6. Improvement in a composite occurs when the score on at least one of the composite questions increases from the pre-test to the post-test.

[Place Table 1 Approximately Here]

Table 2 lists the variables' definitions and summary statistics. Following the approach implemented by USDA, the EFNEP impacts are measured in terms of the percentage of participants showing an improvement for each composite, so consequently these percentages by definition range from 0 to 100. We will refer to the percentages as indices: the FRMP index, the NP index, and the FSP index. On average, the value of the FRMP index is 83% (i.e., 83% of the participants improved in the FRMP composite), the value of the NP index is 88% and the value of the FSP index is 64%. The corresponding standard deviations are FRMP 9%, NP 6%, and FSP 11%.

[Place Table 2 Approximately Here]

Covariate Explanatory Variables

Budget: This variable is defined as the annual budget allocation (in real US $) for each state and US territory. USDA's annual nominal budget allocation for EFNEP is based on the following exogenous rule[1]: 4% is used for federal administration, 10% is equally distributed among the

states, and the remainder is distributed based on states'/territories' population under 125% of income poverty, as determined by the last preceding decennial census. The allocations (2000-2006) were converted into real 2005 dollar values using the Consumer Price Index (CPI). The average amount over all observations (i.e., states/territories and years) is $1,091,583 with a standard deviation of $961,058 (table 2). We hypothesize that controlling for the other covariates, as the budget increases the index scores will improve.

Number of Participants and Participant Characteristics: We include in the analysis the number of adult participants to control for the size of the program. The average number of participants is about 2,880 with a standard deviation of 4,200 (table 2). In order to control for the effects of the youth EFNEP program on the outcomes of the adult program, we also included the number of youth participants in the analysis. The average number of youth participants is about 7,645 with a standard deviation of 9,100.

We include several available socio-demographic categorical characteristic percentages of the participants that are collected by USDA: gender, race/ethnicity, age, educational level, household income, and place of residence as defined in table 2. These socioeconomic variables are typical of those found in the literature on health behavior (e.g., Pollard, et al. 2009). Including these participant characteristics will allow us to determine if, conditional on the amount of money being spent, the program is more effective for one participant characteristic subgroup over another. A finding of differential participant characteristic effects would suggest that targeting certain participant subgroups could lead to a more effective program. When disaggregated by gender, females constitute 89%, and by race, whites and blacks account for about 65% of the participants on average. More than two thirds of the participants (~68%) fall in the 20 to 39 year age range and about 63% have income below 100% of the poverty line.

Unfortunately, about 72% of participants did not provide information on their level of education, but for those who did the largest percentage is for those with education only through some high school (~23%). Less than half of participants (~43%) reside in central cities with populations over 50,000.

Program Characteristics: One may expect different levels of key program characteristics to have different impacts on learning outcomes, consequently the analysis includes information collected by USDA regarding types of lessons and instructors. EFNEP offers four lesson types: (i) individual lessons, (ii) group lessons, (iii) both individual and group lessons, and (iv) other types. A state level analysis by Dollahite and Scott-Pierce (2003) shows that the participants who took individual lessons showed more positive behavior change than those who took group lessons. This seems a very intuitive finding and one that has important implications for how the program should be structured to achieve the largest impact if the result holds at a national level. While individual lessons may be more effective in terms of individual outcomes, they are not necessarily more cost effective. The trade-off that must be considered is the implied decline in effectiveness as students are added to a lesson versus the decrease in the cost per student of the lesson. If the decline in the effectiveness is offset by a greater decline in cost, then group lessons would be more cost effective. The largest percentage of participants attend group lessons (~73%), followed by individual lessons (~21%), group and individual lessons (~6%) and other lesson types (~1%).

Three types of instructors are involved in the program: professionals, paraprofessionals, and volunteers. The professionals train the paraprofessionals and volunteers, and provide technical support. Paraprofessionals are the key personnel responsible for delivering the curriculum. Volunteers primarily assist in program delivery and program management. Being

mostly from the local area, the paraprofessionals and volunteers are expected to be best suited to deliver lessons to local participants. However, if paraprofessionals and volunteers are not as effective as professionals, then this would imply a reallocation of resources toward more professionals may be more effective in achieving the goals of the program. Volunteers constitute the largest percentage of instructors (~85%), followed by paraprofessionals (~13%) and professionals (~2%).

2.4. Empirical Model

The FRMP index, the NP index, and the FSP index are the percentages of individuals improving in the corresponding composite, so by definition, each is restricted to the range 0 to 100% . To account for this restricted range, we use a logistic functional form for each equation (Kmenta 1997). Mathematically, the three equations to estimate have the form

$$(1)\quad Y_i = 100 / (1 + e^{-\alpha_{0i} - \alpha_{1i} \ln B - \boldsymbol{\alpha}_{2i} \boldsymbol{Z}}) + \varepsilon_i \qquad i = FRMP, NP, FSP,$$

where Y denotes the nutrition index (percentage improvement), ln B is the natural log of the budget, $\boldsymbol{Z}$ the vector of control variables, ε the error term with an expected value of zero and the αs are conformable parameters. Equation (1) corresponds to a simple indirect production function.

An important feature of this model is that it is nonlinear and allows for increasing, constant, or decreasing returns to scale in the budget. Specifically, the change in each dependent variable for a one percent change in the budget (i.e., a one unit change in ln B) is the marginal budget effect (MBE), or mathematically from equation (1),

$$(2)\quad MBE_i = \partial Y / \partial \ln B = \alpha_{1i} \times \left[100 \times e^{-\alpha_{0i} - \alpha_{1i} \ln B - \boldsymbol{\alpha}_{2i} Z} / (1 + e^{-\alpha_{0i} - \alpha_{1i} \ln B - \boldsymbol{\alpha}_{2i} Z})^2 \right]$$

The sign of the MBE is determined by the sign of the parameter on the budget, α_{1i}, because the term in brackets is always positive. The magnitude of this effect will vary depending on the value of the budget and other covariates (i.e., the point of evaluation). This formulation will allow for varying returns to scale by state and territory and consequently by region. Furthermore, because the dependent variable is a percentage and the budget is expressed in natural logs, simple calculus reveals that this marginal effect is equal to the ratio of average cost (AC_i) to marginal cost (MC_i) or

$$(3)\quad MBE_i = AC_i / MC_i \qquad i = FRMP, NP, FSP .$$

Basic economics indicates if $AC_i > MC_i$, or $MBE_i > 1$, there are increasing returns to scale, if $AC_i = MC_i$ ($MBE_i = 1$) there are constant returns to scale, and if $AC_i < MC_i$ ($MBE_i < 1$) there are decreasing returns to scale.

Because the error terms across equations are likely to be correlated, we use the non-linear seemingly unrelated regression (NLSUR) method (Wooldridge 2002) for model estimation using the statistical software STATA v11. As many of the explanatory variables are categorical percentages, a reference set of categories must be chosen to avoid perfect collinearity. The reference case here is the percentage of participants under the age of 20, with income less than 50% of the poverty level, education up to the high school level, residing in a central city with a population over 50,000, participated in a group lesson, and instructed by a volunteer. Gender is excluded from the estimation model as ~90% of the sample is female, so the variation is quite limited. Finally, because of repeated observations by states/territories, we use a cluster-robust

covariance matrix as there is likely correlation in the errors over time for the same state/territory (Woolridge 2002).

2.5. Results

The parameter estimates and their corresponding *p*-values are given in table 3 and are only briefly discussed before turning to the returns to scale estimates, as they are the central focus of the analysis. While we recognize the choice of a significance level for inference is somewhat arbitrary, for ease in communication we use a 0.10 significance level cutoff for deciding significance but report p-values such that readers may choose a different cutoff. Following the result presentation is a discussion of the results.

[Place Table 3 Approximately Here]

Budget

Consistent with our hypothesis, the parameter estimate for the budget is positive and statistically significant at conventional levels in the Food Resource Management Practices (FRMP) and the Nutrition Practices (NP) index equations (FRMP: 0.15, *p*-value = 0.01 and NP: 0.12, *p*-value = 0.04). The parameter estimate for the budget is not significantly different from zero in the Food Safety Practices (FSP) index equation (0.00, *p*-value = 0.98).

Other Covariates

Very few of the participant characteristics are significant. We focus on those that are significant in more than one equation. As the percentage of participants with incomes over the 150% of the poverty line increases, the NP index (0.06, *p*-value =0.01) and FSP index (0.03, *p*-

value = 0.09) both increase. Though still positive, this effect is not statistically significant in the FRMP index (0.04, *p*-value = 0.11). When the percentage of participants who did not report their age increases relative to the percentage of younger participants (i.e., those of age less than 20), all indices decrease (FRMP: –0.02, *p*-value = 0.07; NP: –0.02, *p*-value = 0.04: FSP: –0.02, *p*-value = 0.08).The other control variable that is statistically significant in more than one equation is the place of residence variable (rural non-farm). Relative to the base place of residence (central cities), rural non-farm scored significantly higher on all three indices (FRMP: 0.01, *p*-value = <.001; NP: 0.01, *p*-value = 0.001: FSP: 0.01, *p*-value = <.001).

With respect to the program characteristics, there is no apparent significant difference in the index scores associated with the lesson types, with exception of the "other" lesson type (Lesson4) in the FSP index equation. As the percentage of participants that did not participate in a group and/or individual lesson increases the FSP index score decreases (–0.01, *p*-value = 0.06). The variable percentage of professional instructors is significant in NP and FSP index equations. An increase in the percentage of professional instructors increases both the NP index (0.03, *p*-value = 0.09) and the FSP index (0.02, *p*-value = 0.08).

Returns to Scale

As indicated above, the marginal budget effect is nonlinear and varies by evaluation point. The marginal budget effect for each index is evaluated at the average value of all covariates at the (i) national level (all states/territories), at the average value for all covariates (ii) for the corresponding regional[2] level, and at the average value of all covariates (iii) for the corresponding state/territory level. The variance and associated *p*-values for these marginal budget effects are based on the delta method and account for clustering (Wooldridge 2002). The

relevant hypotheses tested are for decreasing, constant, and increasing returns to scale, which amounts to testing if the marginal budget effects are less than one, one, and greater than one respectively, and so these are the results discussed. To avoid redundancy, we only discuss the returns to scale for the FRMP and NP indices because the marginal budget effect for the food safety practices (FSP) index is not significantly different from zero, regardless of level of aggregation (national, regional, state), implying decreasing returns to scale at all levels. Though the emphasis in this section is on statistical significance, the final section incorporates a broader analytical perspective by also discussing the economic significance of the results (McCloskey 1985).

At the national level, the marginal budget effect for the FRMP index is 2.01 and for the NP index is 1.20. Though the FRMP marginal budget effect is 2.01, the returns to scales test indicates it is not statistically different from one at the 0.10 significance level, though it is greater than one (increasing returns to scale) at the 0.11 level (*p*-value = 0.11). Similarly, the marginal budget effect of 1.20 for NP is also not statistically different from one, indicating constant returns to scale.

There is however a great deal of variability by region and state/territory in the marginal budget effects and hence returns to scale. Figures 1 and 2 provide the marginal budget effects evaluated at the regional level and at the state/territory level for the FRMP index and the NP index, respectively.

[Place Figure 1 Approximately Here]

For the FRMP index (Figure 1), the marginal budget effect is greater than 1.00 for all regions. The highest returns to scale are in the 'West' ($MBE_{FRMP} = 2.30$), followed by the

'Midwest' ($MBE_{FRMP} = 2.15$), the 'Northeast' ($MBE_{FRMP} = 2.13$), the 'South' ($MBE_{FRMP} = 1.76$), and the 'Territories' ($MBE_{FRMP} = 1.08$). The returns to scale tests indicate that three of the five regions: the 'West' (*p*-value = 0.08); the 'Midwest' (*p*-value = 0.09); and the 'Northeast' (*p*-value = 0.09) have statistically increasing returns to scale, and the two regions 'South ' and 'Territories' have statistically constant returns to scale. Yet, even within each region, there is some variability across states. With exception of the Territories, each region has some states that have statistically significant increasing returns to scale in the FRMP index. Seven of 13 states in the West had statistically increasing returns to scale – WA (2.31), MT (2.71), AZ (2.51), CO (2.37), NM (2.24), AK (2.27), and OR (2.17). Seven of the 12 states in the Midwest had statistically increasing returns to scale – SD (2.49), KS (2.47), IA (2.17), OH (2.83), MN (2.18), ND (2.20), and IL (2.08). Five of the 9 states in the Northeast had statistically significant increasing returns to scale – MA (2.07), RI (2.76), CT (2.28), ME (2.16), and NJ (2.43). Only four of the 16 states in the South had statistically increasing returns to scale – GA (2.67), TN (1.99), VA (2.15), and FL (2.15). All other states have statistically constant returns to scale in the FRMP index.

[Place Figure 2 Approximately Here]

For the NP index (Figure 2), the marginal budget effect is numerically greater than 1.00 for four out of the five regions. The highest numerical returns to scale are in the 'Midwest' ($MBE_{NP} = 1.32$), followed by the 'West' ($MBE_{NP} = 1.30$), the 'Northeast' ($MBE_{NP} = 1.19$), the 'South' ($MBE_{NP} = 1.10$), and the 'Territories' ($MBE_{NP} = 0.63$), but the statistical tests indicate all regions have statistically constant returns to scale in the NP index. In contrast to the FRMP index, though there is some numerical variability across states within each region, all states have

statistically constant returns to scale, with the exception of American Samoa and Northern Marianas, who have decreasing returns to scale in the NP index.

2.6. Discussion and Implications

The participant characteristics are measured in terms of categorical percentages (percentages of the sample falling within a category) so the effects (parameter estimates) associated with these variables are effectively measuring differences due to shifting the distribution of individuals on the characteristic domain. The closest literature to our analysis is on child education attainment (e.g., Gunn et al., 2000; Rumberger and Palardy, 2005). In these studies the main demographic variables are those of the parents so any comparisons with our results would exceed the bounds of valid comparison. Furthermore, in these child education attainment studies monetary inputs are usually NOT included as an explanatory variable, so by standard omitted variable analysis the effects of parent or student characteristics include an omitted variable bias, which may lead to their significance. We explored this possibility and estimated all models *without* the budget variable and found that the same variables were still insignificant. Alternatively one could argue that the statistical insignificance of many of the explanatory variables in our model could be due to the categorical nature of the variables. It is well established that the categorical variables can mask the underlying significant relationships if the variable is better measured by a continuous variables, such as income (e.g., MacCallum, et al. 2002).

Keeping in mind these measurement caveats, the results of this analysis suggest there are *generally* no major inherent participant characteristic biases affecting the effectiveness of the

EFNEP. In addition, there does not appear to be any general major program characteristic that is more effective than any other characteristic. These findings are actually useful and positive as they suggest the program appears to be uniformly effective across these different personal and program characteristics as a whole, so efforts to change the personal and program characteristic profiles would be predicted not to have a significant effect on the effectiveness of the program. However, our results show that states with higher proportion of participants from rural areas generally tend to perform better than the states with higher proportion of participants from the urban areas. What this implies is that other differences across states and regions also should be investigated to help explain differences in outcomes not explained by these factors, such as geographic, cultural, and other economic based differences.

For two of the three indices used by USDA to measure the success of the adult EFNEP, the FRMP and the NP, the money being spent has a positive and significant effect. Three out of five regions show statistically significant increasing returns to scale in the FRMP. Further, 23 states show statistically significant increasing returns to scale in the FRMP index: AK, AZ, CO, CT, FL, GA, IL, IA, KS, ME, MA, MN, MT, NJ, NM, ND, OH, OR, RI, SD, TN, VA, and WA. For the NP, almost all states and regions show constant returns to scale.

McCloskey (1985) makes a compelling case that often too much emphasis is placed on statistical significance and not enough on economic significance. She gives a vivid example of testing purchasing power parity, which boils down to a test of $\beta = 1$ in a model (pp. 201-202). Just because someone has a very large sample and rejects the estimate that $\beta = 0.999$ is equal to one, from an economic significance or common sense perspective, purchasing power parity would not be rejected. It should be clear that her general argument would apply to tests of returns to scale and hence are very relevant here as well. Though from a statistical perspective

the majority of states have constant returns to scale in the FRMP and NP indices, in terms of economic significance, more latitude should be applied in interpreting the numbers. For the FRMP, there are 29 states with return to scale estimates over 2.00 and another 18 with return to scale estimates between 1.50 and 2.00. For the NP, there are 42 states with return to scale estimates over 1.00. All of this suggests that, given the stated objectives and outcome indices for the EFNEP utilized by USDA, the program is very effective within a broader evaluation perspective.

How can these results be used to improve the EFNEP? Much of the issue of improving the EFNEP comes down to first having objective valid information on effectiveness at the state level. Though a lot of information is exchanged between EFNEP administrators, until we actually know which states are doing better relative to others it will remain unclear who has the more effective program and hence more effective information. The model and returns to scale estimates here provide this first required piece of information. Much of the usefulness of this research is then in simply identifying outcome areas and states that are doing well relative to others and all that may be required is to share information on different practices, which can easily be done at the National EFNEP Conference each year.

The results shed light on two main areas where the EFNEP administrators and stakeholders should look to make improvements. First, it is clear that the money being spent on food safety is not having much of an impact on the food safety practices (FSP) index. Revaluating the general EFNEP approach to food safety practices and considering alternative approaches may actually improve the food safety outcomes associated with EFNEP. Second, even for the food resource management practices (FRMP) and nutrition practices (NP) there is a great deal of variability even within regions on returns to scale. EFNEP administrators and

stakeholders in a relatively low performing state need only look to the higher performing states in their region for insights into how their program operations differ and for suggestions for ways of improving their program operations.

Regarding limitations, two seem especially noteworthy. First, behavioral checklists and categorical based surveys, such as that used by USDA, are utilized mainly for their high response rate and practicality in implementation. They are easy targets for criticism as they must navigate the often unclear tradeoffs between validity, reliability, response burden, and practicality in implementation. The procedures implemented by USDA in developing the behavior checklist are in line with the recommendations of the literature (e.g., Contento, et al. 2002; Kristal, et al. 1990) and there is some evidence that behavior checklist questions can correlate reasonably well with behavioral and some biological changes (e.g., Murphy, et al. 2001). However, when there is interest in possibly multiple domains or constructs, the improvement in a particular domain could simply depend on the number or type of questions asked, as pointed out by a reviewer. For example, the 10 item behavioral checklist used by USDA contains only two questions for the Food Safety Practices (FSP) compared to five questions for the Nutrition Practices (NP) index. This implies that even if questions are answered randomly, the probability of improving on at least one questions, as considered by USDA, will be higher the more questions that are asked on a given domain. This is certainly an area where more research is needed on the quality of the USDA behavioral checklist and related indicators. Until this time, these existing national indicators are the obvious place to start in any type of national level analysis and discussion of returns to scale and costs of the EFNEP.

Second, the stated objectives of the EFNEP focus on education and behavior and it is well known that improved knowledge and behavior may not translate into health benefits. A few

small state level EFNEP cost-benefit studies do indicate the health benefits, in terms of reduced health spending, exceed the costs (Dollahite, Kenkle, and Thompson 2008; Joy, Pradhan, and Goldman 2006; Rajgopal, et al. 2002; Schuster, et al. 2003). At the national level this is a daunting question. However, regardless of the theory used to support the notion of changing health through an education program, a central component is to change behavior and this is one of the main direct goals stated by EFNEP. Consequently, the type of analysis conducted here seems the appropriate starting point for any returns to scale and cost analysis. As usual, these two points imply further research is needed but, at this time, the best available information indicates the EFNEP is an effective program in achieving its stated goals.

Footnotes

[1] This rule was revised in 2008, but is relevant for the time for which the data are used in this study.

[2] We divided the US into four regions based on the regional divisions used by the United States Census Bureau. For completeness of the analysis, all US-territories were categorized into a fifth region.

References

Anliker, J., W. Willis, and S. Montgomery. "The Development and Testing of the Behavior Checklist Questions for the EFNEP Evaluation/Reporting System." Internet site: http://www.csrees.usda.gov/nea/food/efnep/ers/documentation/checklist-development.pdf (Accessed August 12, 2011).

Arnold, C. G., and J. Sobal. "Food Practices and Nutrition Knowledge After Graduation From The Expanded Food And Nutrition Education Program (EFNEP)." *Journal of Nutrition.*32(2000):130-138.

Contento, I. R., J. S. Randell, and C. E. Basch. "Review and Analysis of Evaluation Measures Used in Nutrition Education Intervention Research. *Journal of Nutrition Education and Behavior*. 34 (2002):2-25.

Dickin, K. L., J. S. Dollahite, and J. P. Habicht. "Nutrition Behavior Change Among EFNEP Participants Is Higher At Sites That Are Well Managed And Whose Front-Line Educators Value The Program." *Journal of Nutrition*. 135(2005):2199–2205.

Dollahite, J. S., and M. Scott-Pierce. "Outcomes of Individuals vs. Group Instruction in EFNEP." *Journal of Extension*. 41 (2003). Internet site: http://www.joe.org/joe/2003april/a4.php (Accessed August 11, 2011).

Dollahite, J. S., D. Kenkel, and C.S. Thompson. "An Economic Evaluation of The Expanded Food and Nutrition Education Program." *Journal of Nutrition Education and Behavior*. 40 (2008):134-43.

General Accounting Office (GAO). *Nutrition Education*: *USDA Provides Services through Multiple Programs, but Stronger Linkages Among Efforts are Needed.* Washington, DC: United States General Accounting Office, 2004.

Gunn, J.B., P.K. Klebanov, and G.J.Duncan. "Ethnic Differences in Children's Intelligence Test Scores: Role of Economic Deprivation, Home Environment, and Maternal Characteristics." *Child Development*. 67 (1996): 396-408.

Joy, A. B., V. Pradhan, and G. Goldman. "Cost-Benefit Analysis Conducted For Nutrition Education in California" *California Agriculture*. 60 (2006):185-191.

Kmenta, J. *Elements of Econometrics*. 2nd Edition. Ann Arbor: University of Michigan, 1997.

Kristal, A. R., B. F. Abrams, M. D. Thornquist, L. Disogra, R. T. Croyle, A. L. Shattuck, and H. J. Henry. "Development and Validation of a Food Use Checklist for Evaluation of Community Nutrition Interventions." *American Journal of Public Health*. 80 (1990):1318-1322.

Marmot, M. and M. Wadsworth, eds. *Social Determinants of Health*. Oxford: Oxford University Press, 1999.

MacCallum, C. R., S. Zhang, K.J. Preacher, and D.D. Rucker. "On the Practice of Dichotomization of Quantitative Variables." *Psychological Methods*. 7 (2002): 19-40.

McCloskey, D. N. "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests." *American Economic Review*. 75 (1985): 201-205.

Murphy, A. P., L. L. Kaiser, M. S. Townsend, and L. H. Allen. "Evaluation of Validity of Items for a Food Behavior Checklist." *Journal of American Dietetics Association*. 101 (2001):751-761.

Pollard, C., M. Miller, R. Woodman R, R. Meng, and C Binns. "Changes in Knowledge, Beliefs, and Behaviors Related to Fruit and Vegetable Consumption Among Western Australian Adults From 1995 to 2004." *American Journal of Public Health*. 99 (2009):355-361.

Rajgopal R., R. H. Cox, M. Lambur, and E.C. Lewis. "Cost-Benefit Analysis Indicates the Positive Economic Benefits of the Expanded Food and Nutrition Education Program

Related to Chronic Disease Prevention." *Journal of Nutrition Education Behavior*. 34 (2002):26-37.

Rumberger, R.W. and G.J.Palardy. "Test Scores, Dropout Rates, and Transfer Rates as Alternative Indicators of High School Performance." *American Educational Research Journal*. 42 (2005): 3-42.

Shephard, RW. *Indirect Production Functions*. Verlag Anton Hain, Meisenheim Am Glad, 1974.

Schuster, E., Z. L. Zimmerman, M. Engle, J. Smiley, E. Syversen, and J. Murray. "Investing in Oregon's Expanded Food and Nutrition Education Program (EFNEP): Documenting Costs and Benefits." *Journal of Nutrition and Education Behavior*.35 (2003):200-206.

United States Department of Agriculture (USDA)- National Institute of Food and Agriculture. *Expanded Food and Nutrition Education Program*. Internet site: http://www.nifa.usda.gov/nea/food/efnep/efnep.html (Accessed August 10, 2009a).

_____ National Institute of Food and Agriculture (NIFA). *NIFA Budget Information. Archived Extension Activities.* Internet site: http://www.csrees.usda.gov/about/offices/budget.html (Accessed August 11, 2009b).

_____ Cooperative State Research, Education, and Extension Service. *Nutrition Education Evaluation and Reporting System*. Internet site: http://www.csrees.usda.gov/nea/food/efnep/neers5/neers5.html (Accessed August 11, 2009c).

Wooldridge, J. M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts. The MIT Press, 2002.

Table 1. Behavior Checklist Question Components of Nutrition Knowledge Indices

	Behavior Checklist Questions	Indices
1	How often do you plan meals ahead of time?	FRMP, NP
2	How often do you compare prices before you buy food?	FRMP
3	How often do you run out of food before the end of the month?	FRMP
4	How often do you shop with a grocery list?	FRMP
5	How often do you let these foods (meat and dairy) sit out for more than 2 hours?	FSP
6	How often do you thaw frozen foods at room temperature?	FSP
7	When deciding what to feed your family, how often do you think about healthy food choices?	NP
8	How often have you prepared foods without adding salt?	NP
9	How often do you use "Nutrition Facts" on the label to make food choices?	NP
10	How often do your children eat something in the morning within two hours of waking up?	NP

Note: FRMP – Food Resource Management Practices; FSP – Food Safety Practices; NP – Nutrition Practices.

Table 2. Variable Descriptions and Summary Statistics

Variable	Description	Mean(SD)
Dependent Variables		
FRMP index	Percentage of participants who improved in one or more Food Resource Management questions	82.72(8.62)
NP index	Percentage of participants who improved in one or more Nutritional Practice questions	87.91(5.85)
FSP index	Percentage of participants who improved in one or more Food Safety Practice questions	64.44(11.40)
Covariates		
Budget	Annual federal budget allocation per state/territory (in real dollar units)	1,091,583 (961,058.10)
Participant	Number of participants in a given state for a given year (1000s)	2.88(4.20)
Youths	Number of youth participants in a given state for a given year (1000s)	7.64 (9.10)
Female	Percentage of female participants	89.29(8.20)
Male*	Percentage of male participants	10.70(8.21)
White	Percentage of White participants	40.34(26.21)
Black	Percentage of Black participants	25.02(25.79)
Hispanic	Percentage of Hispanic participants	19.65(23.01)
Other race*	Percentage of participants from other race	14.97(27.30)
AgeUnder20*	Percentage of participants with age under 20 years	13.11(7.54)

Table 2. Variable Descriptions and Summary Statistics (con't)

Age20-29	Percentage of participants with age 20 to 29 years	45.56(18.34)
Age30-39	Percentage of participants with age 30 to 39 years	22.71(11.18)
Age40-49	Percentage of participants with age 40 to 49 years	10.58(6.67)
Age50-59	Percentage of participants with age 50 to 59 years	3.62(3.52)
Age60Plus	Percentage of participants with age above 60 years	2.26(4.51)
AgeNA	Percentage of participants without age information	2.16(3.79)
Income50% of poverty level*	Percentage of participants whose household income is less than or equal to 50% of poverty level	36.75(19.22)
Income100% of poverty level	Percentage of participants whose household income is between 51 to 100% of poverty level	25.69(10.98)
Income150% of poverty level	Percentage of participants whose household income is between 101 to 150% of poverty level	8.82(5.96)
Income150Plus% of poverty level	Percentage of participants whose household income is more than 150% of poverty level	4.34(4.85)
IncomeNA	Percentage of participants for whom household income information is not available	24.39(22.37)
EduNA	Percentage of participants without information on education status	71.82(36.99)
Highschool*	Percentage of participants with education up to high school	22.67(30.21)
Some college	Percentage of participants with some college level education	4.56(6.94)

Table 2. Variable Descriptions and Summary Statistics (con't)

College grad and up	Percentage of participants who graduated from college or above	1.70(2.92)
Rural farm	Percentage of participants who reside in rural farm area	2.98(8.70)
Rural non- farm	Percentage of participants who reside in town with population under 10,000 and rural non-farm area	26.82(24.89)
Small town	Percentage of participants who reside in town and cities with population 10,000 to 50,000 and their suburbs	21.34(16.04)
Suburb of cities	Percentage of participants who reside in suburb of cities with population over 50,000	5.34(6.51)
Central Cities*	Percentage of participants who reside in central cities with population above 50,000	43.52(29.07)
Group lesson*	Percentage of participants who were delivered group lessons	72.70(23.25)
Individual lesson	Percentage of participants who were delivered individual lessons	20.53(18.63)
Both group and individual lesson	Percentage of participants who were delivered both group and individual lessons	5.59(8.20)
Other lesson	Percentage of participants for whom other types of instruction were used	1.17(3.69)

Table 2. Variable Descriptions and Summary Statistics (con't)

Professionals	Percentage of professionals among the instructors	2.05(4.91)
Paraprofessionals	Percentage of para-professionals among the instructors	12.67(17.88)
Volunteers*	Percentage of volunteers among the instructors	85.27(19.77)

Note: * variables used as base in the model. Mean is calculated as the overall average across states and territories, for all seven years.

Table 3. Nonlinear Seemingly Unrelated Regression Parameter Estimates

	System Equation Results					
	FRMP		NP		FSP	
Variable	Parameter	p-value	Parameter	p-value	Parameter	p-value
Log of Budget	0.146	0.01**	0.116	0.04**	0.001	0.98
Number of participants	0.015	0.35	0.012	0.42	0.003	0.84
Number of youth participants	0.000	0.99	-0.005	0.47	0.003	0.68
Income100% of poverty level	-0.011	0.17	-0.008	0.27	-0.001	0.94
Income150% of poverty level	-0.034	0.16	-0.030	0.15	-0.027	0.12
Income150Plus% of poverty level	0.039	0.11	0.058	0.01**	0.026	0.09*
Income NA	-0.002	0.50	0.000	0.87	0.001	0.82
White	-0.003	0.52	0.000	0.95	0.003	0.39
Black	-0.005	0.28	-0.002	0.58	0.003	0.41
Hispanic	0.004	0.25	0.003	0.45	0.010	0.02**
Age20-29	-0.005	0.46	0.000	0.98	0.002	0.71
Age30-39	0.012	0.29	0.015	0.20	0.027	0.01**
Age40-49	-0.008	0.67	0.000	0.99	-0.009	0.54
Age50-59	-0.109	0.25	-0.031	0.42	-0.071	0.20
Age60Plus	0.111	0.19	0.025	0.32	0.030	0.40
Age NA	-0.023	0.07*	-0.021	0.04**	-0.017	0.08*
Some college	-0.005	0.80	0.002	0.90	-0.014	0.22
College grad and up	-0.014	0.71	-0.015	0.67	-0.001	0.97

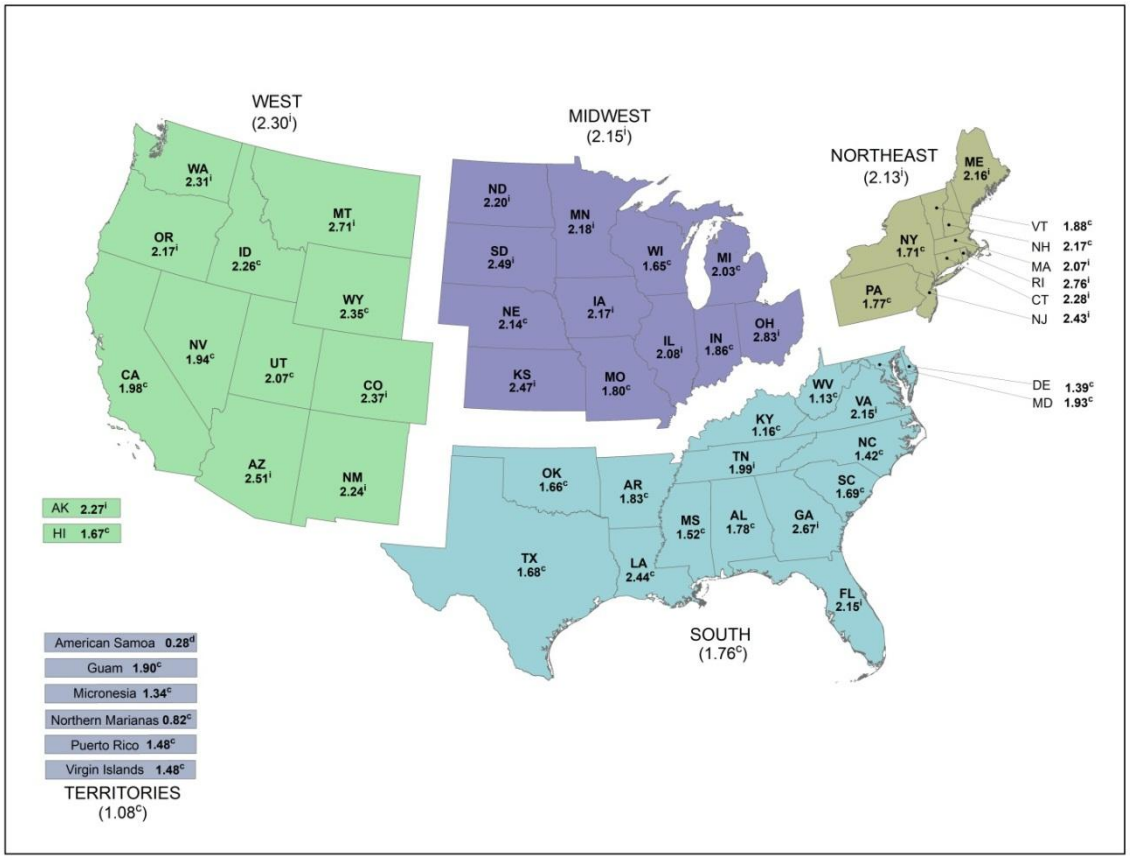
Table 3. Nonlinear Seemingly Unrelated Regression Parameter Estimates (con't)

Education NA	-0.003	0.48	0.000	0.89	-0.005	0.07*
Rural farm	0.003	0.84	0.011	0.52	-0.017	0.16
Rural non-farm	0.012	0.00**	0.010	0.00**	0.010	0.00**
Small town	0.006	0.22	0.006	0.19	0.004	0.29
Suburb of cities	-0.005	0.41	-0.015	0.08*	-0.002	0.74
Individual lesson	0.006	0.12	0.005	0.18	0.000	0.90
Both group and individual lesson	0.007	0.45	-0.005	0.47	0.004	0.64
Other lesson	-0.010	0.34	-0.003	0.82	-0.014	0.06*
Professionals	0.033	0.15	0.035	0.09*	0.024	0.08*
Paraprofessionals	0.006	0.25	0.002	0.74	0.002	0.46

Note: ** Significant at 5%, * Significant at 10%

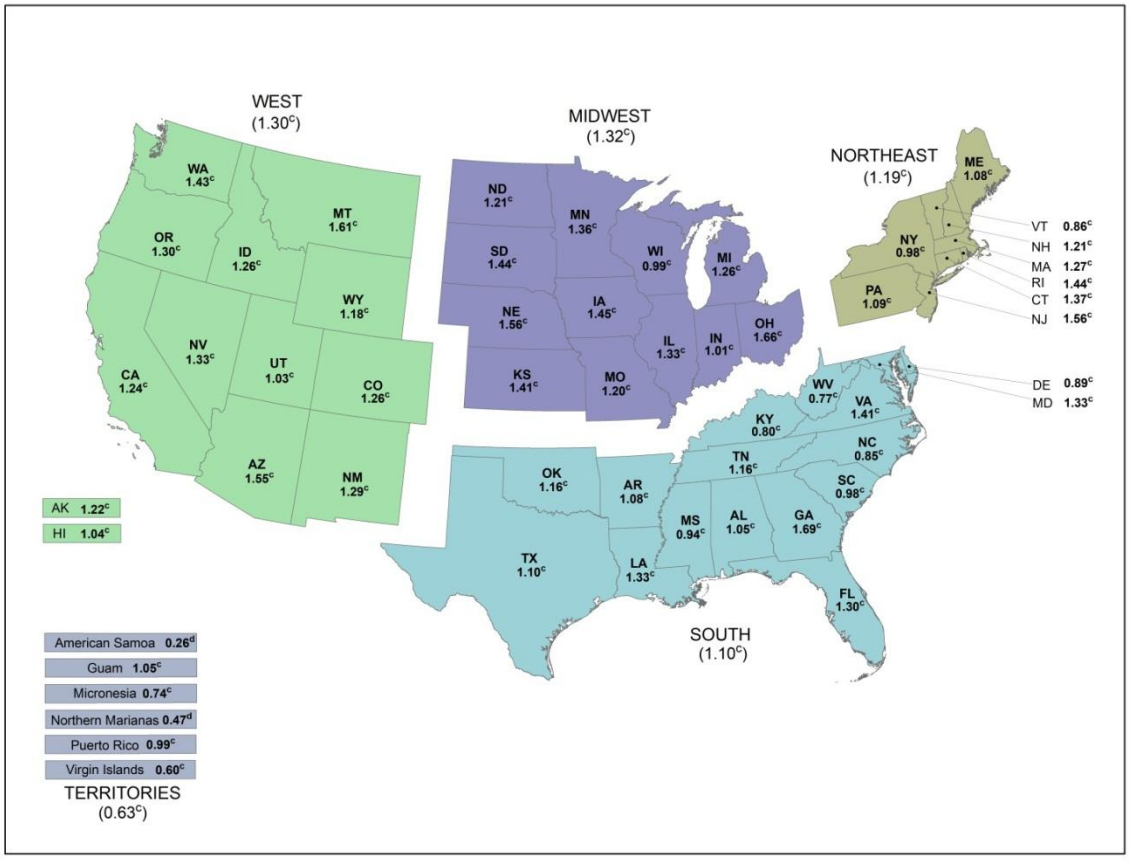
Goodness of fit statistics is given by the correlation coefficient between the predicted values and the actual values: FRMP (0.46), NP (0.48), and FRMP (0.56).

Figure 1. Returns to Scale on Food Resource Management Practices by State/Territory



Note: Statistical test indicating d = decreasing returns to scale; c= constant returns to scale; i = increasing returns to scale at 0.10 significance level.

Figure 2. Returns to Scale on Nutrition Practices by State/Territory



Note: Statistical test indicating d = decreasing returns to scale; c= constant returns to scale; i = increasing returns to scale at 0.10 significance level.

Chapter 3: Cost Effectiveness Analysis of Youth Expanded Food and Nutrition Education Program in Virginia

3.1. Introduction

3.1.1. Problem Statement and Significance of the Study

The Expanded Food and Nutrition Education Program (EFNEP) is one of the largest federally funded nutrition education programs in the United States and is administered by the United States Department of Agriculture (USDA) (General Accounting Office 2004). The aim of EFNEP is to serve limited resource youths and adults by helping them acquire "the knowledge, skills, attitudes, and changed behavior necessary for nutritionally sound diets, and to contribute to their personal development and improvement of the overall family diet and nutritional well-being" (USDA 2013). In operation for more than 40 years and now in all 50 states and 6 territories, the EFNEP has become a cornerstone in US nutrition education (USDA 2013).

EFNEP appropriates more than $66 million dollars of the federal money, every year. Given the tight budgetary conditions, there has been an increasing pressure for the evaluation of federally funded programs to improve their effectiveness and to ensure program accountability (General Accounting Office [GAO] 2004). While several research studies have examined and documented the effectiveness, and cost benefits for *adult* EFNEP programs (Rajgopal et al. 2002; Joy et al. 2006; Wessman and Jensen 2002; Schuster et al. 2003; Dollahite et al. 2008), there is a dearth of research on the youth program, which constitutes more than 75% of program beneficiaries (more than 500,000 youth participants in year 2011). Townsend and colleagues (2006) are the only scholars to study and document the effectiveness of youth EFNEP to date. No cost effectiveness analyses of youth EFNEP programs exist.

Part of the reason for the dearth of evaluation studies on youth EFNEP has been the lack of comprehensive, valid, and reliable assessment tools to measure program effectiveness (USDA 2008). Further, unlike the adult EFNEP, no federal guideline mandates the use of any specific survey questions/instruments, or even curricula for youth EFNEP. For evaluation, the state programs often create their own survey instruments. This results in the lack of consistency and standardization across youth EFNEP programs and evaluation.

To address the challenges in evaluating the youth EFNEP programs, this project envisioned developing a cost effectiveness analysis (CEA) model for the youth EFNEP program in Virginia. A valid and reliable instrument to measure the program outcomes was developed and used in conducting the CEA. Results are presented in terms of the cost effectiveness ratios (CER) which expresses the amount of resources spent (in $) by the youth EFNEP program in producing the outcomes as dictated by the program goals. The project was guided, and closely monitored, by an advisory board consisting of nationally recognized faculty with expertise in nutrition, EFNEP, economics, item response theory, and program evaluation. Inputs from subject matter experts were instrumental to develop the perspective for the study and to enrich the validity and usability of this study.

The economic cost of a program in producing an impact is a basic question from a policy perspective. At times of economic pressure, education programs are easy targets for reducing the government spending. In such times, estimates of the CER could assist policy makers to prioritize the allocation of scarce resources. This study provides a concrete application and example of conducting a CEA for youth EFNEP (or other nutrition education programs) that other states/programs can consider applying to their own programs. EFNEP stakeholders could

examine their programs utilizing this model in order to ensure that EFNEP youth nutrition programs are cost effective.

3.1.2. Goal and Objectives

The ultimate goal of this chapter is to create a cost effectiveness modeling procedure so that state and national level Extension faculty, administrators, and researchers can calculate and communicate the cost effectiveness of youth EFNEP programs (as well as other youth nutrition education programs).

This study demonstrates the use of the cost effectiveness model with a case study of the Virginia youth EFNEP. In this context, the specific objectives of this study are

- To measure costs of implementing youth EFNEP program in Virginia
- To measure effects of youth EFNEP program in Virginia
- To estimate the cost effectiveness ratio of the youth EFNEP in Virginia

3.1.3. Description of Virginia Youth EFNEP: Healthy Weights for Healthy Kids Program

The core of the youth EFNEP program in Virginia is the Healthy Weights for Healthy Kids (HWHK) curriculum, which is the foundation of this study.

Based on the Dietary Guidelines for Americans, the HWHK curriculum was developed by Virginia Cooperative Extension [VCE] as a response to the childhood obesity epidemic and is targeted to the youths aged between 7 and 14 years (Serrano et al. 2007). The HWHK curriculum is based on the experiential learning model and contains the following six lessons with different activities possible for each lesson:

1. Smart Foods – Children use MyPlate to learn about nutrition and to make smart food choices;

2. Smart Choices – Children explore different ways to enjoy eating, focusing on "quality" over "quantity," by measuring portion sizes of different meals and food products, planning a healthy plate or meal, eating slowly, and using different senses when eating;

3. Smart Drinks – Children investigate the amount of sugar and fat in popular beverages and can even learn how to make a nutritious homemade soda;

4. Smart Snacks – Children have the option of using the food label to compare snacks, making healthy "snackwiches," exploring what puts "whole" into whole grains, or creating advertising campaigns for snacks;

5. Smart Activities – Children get an opportunity to have fun and move in this lesson. They also can find out about calories and what inventions and discoveries have taken place that limit our country's need to be active; and

6. Smart Image – Children view, reflect, and discuss different media images and societal attitudes toward body size to improve attitudes and respect toward different sized and shaped individuals and to focus on "what is inside."

The curriculum is designed to be delivered in six one hour lessons, and cover activities from each of the six lessons. The curriculum was pilot-tested by EFNEP and the Supplemental Nutrition Assistance Program – Education Program (SNAP-Ed) (formerly Food Stamp Nutrition Education Program) educators and peer-reviewed before adoption.

The youth EFNEP program is delivered by para-professionals called *program assistants* (PAs). The PAs manage and deliver the HWHK curriculum to the participants in their designated counties. First, PAs select the schools that have at least 50% of students eligible for free or reduced school lunch. In collaboration with local schoolteachers, PAs then identify the group of students' to participate in the program, and deliver the program. The curriculum is provided during school (considered "enrichment), in after-school care programs and through 4-H EFNEP clubs, day camps, residential camps, community centers, neighborhood groups, and home gardening workshops. Given the significant heterogeneity among youth participants (for example, differences in geographic location (rural and urban), and age range), the delivery of programs takes on various forms. The PAs, local to the community in which they serve, decide the content from within the lessons, order the content, and determine the frequency of meeting for class for each group of participants.

3.2. Methods

3.2.1. Cost Effectiveness Analysis

Cost effectiveness analysis is a common method used to evaluate and to compare the effects and costs of programs designed to improve health (Gold et al. 1996). Cost effectiveness analysis (CEA) adopted in this study is an alternative to the cost benefit analysis (CBA), commonly adopted for evaluating the adult EFNEP program. In CBA, benefits of health intervention are expressed in dollar terms rather than in terms of a nonmonetary effectiveness measure in CEA. Although the scope of application of CBA is broader than that of CEA, the CEA is also informative in stating the program success in achieving its goal. In fact, CEA are

favored over CBA for evaluating health programs as it avoids the ethical concerns over converting the program benefits into monetary units (Gold et al. 1996).

The CEA describes a program in terms of the (incremental) cost effectiveness ratio (CER), which is expressed as

$$\text{CER} = \frac{\text{Total Cost}_{\text{A}} - \text{Total Cost}_{\text{B}}}{\text{Total Effects}_{\text{A}} - \text{Total Effects}_{\text{B}}}$$

where, A and B could be two different programs being compared. When the program is compared with no program (status quo), as is the case in this study, then the formula becomes

$$\text{CER} = \frac{\text{Total Cost}}{\text{Total Effects}} \qquad (1)$$

Equation (1) is the key formula used in this analysis. The numerator, total cost, refers to the dollar value of inputs used to design, deliver, and to maintain the program. The denominator, total effects, may be the targeted final outcomes of the program or could be intermediate outcomes (Gold et al. 1996) specific to the program's goal.

When estimating the CER using data from multiple units (such as different counties or programs), the estimates can be different depending upon the level of aggregation and averaging done in costs and effects data that comes from various units. In this chapter, CER derived by using equation 1 are expressed as total CER and average CER. The distinction between the two measures of CER is important at the outset to avoid confusion. Let TCER be the total CER, where the numerator is sum of all cost and the denominator is sum of all effects, over all counties. Let ACER be the average CER across counties, that is, the average of CER's estimated separately for each county. Let, n be the number of programs/counties included in the study, and

$C_1, C_2, \ldots, C_n$ be the total cost incurred for programs/counties $1, \ldots, n$. Similarly, $E_1, E_2, \ldots, E_n$ be the effects of each county. By definition,

$TCER = \frac{\sum_i C_i}{\sum_i E_i}$, and $ACER = \frac{1}{N}\sum_i \frac{C_i}{E_i}$.

Empirically, the estimates of TCER and ACER derived using the same data can either be greater than, equal to, or less than each other. In an attempt to disentangle the relationship between the two estimators, we proceed as follows:

$TCER \lessgtr ACER$

Or, $\frac{\sum_i C_i}{\sum_i E_i} \lessgtr \frac{1}{N}\sum_i \frac{C_i}{E_i}$

Some algebraic manipulations establish the following relationship

$0 \lessgtr \sum_i \frac{C_i}{E_i}(\bar{E} - E_i)$ $[\because \bar{E} = \frac{\sum_i E_i}{N}]$

The relationship between TCER and ACER is determined by the above expression. Because $C_i \geq 0$ and $E_i \geq 0 \; \forall \; i$, the ratio $\frac{C_i}{E_i}$ is always positive. The relationship between TCER and ACER thus depends on terms inside the parenthesis only, that is the difference between the average effects across all counties to effects of county i. The estimates of TCER and ACER will be equal only when $E_i = \bar{E} \; \forall \; i$. Whether TCER < ACER or TCER > ACER cannot be pre-determined because it depends on whether $E_i > \bar{E}$, or $E_i < \bar{E}$. From basic mathematics, we know that E_i cannot be always greater than or less than $\bar{E} \; \forall \; i$. Even in cases where E_i is greater than or less than $\bar{E}$ for some i, the total sum of the product of the ratio $\frac{C_i}{E_i}$ and the mean deviation

of effects can be greater or less than zero rendering the estimates of either TCER or ACER to be greater or less than each other.

In addition to the differences in TCER and ACER, the estimates of the CER[1] and its interpretations vary widely with the measures of costs and effects. In order to ensure the comparability of the CER across similar programs standard procedures to measure costs and effects were developed and used in this study.

3.2.2. Conceptualization of Costs and Effectiveness Measures

This project envisioned to develop a cost effectiveness model for the youth EFNEP program that could be used by youth EFNEP programs across the US and also by other similar programs. The whole project cycle was closely monitored and guided by the "advisory board" that consists of nationally recognized faculty with expertise in nutrition, EFNEP, economics, item response theory, and program evaluation. This is a popular mechanism for assuring content validity in the design of the instruments. A brief background on the formation of the advisory committee, and the process in which the project evolved provides a context for the discussion of the development of the instruments developed.

3.2.2.1. Formation of Advisory Committee and Organization of Workshop

A 2-day expert panel workshop was conducted at the beginning of the project, 11-12 May 2009, at Virginia Tech, Blacksburg, VA. The panel members consisted of 12 State and National experts who were representative of economics, nutrition, behavioral science, EFNEP, extension, and evaluation within the Extension context. Their inputs were instrumental in conceptualizing

[1] The term CER is used to refer both TCER and ACER, unless specifically mentioned.
[2] A separate review article summarizing these instruments have been published (Hernandez-Garbanzo et al. 2013).

the project, developing instruments, interpreting findings, and guiding the different cost effectiveness approaches.

The main goal of the workshop was to gather and document opinions from these experts in order to create a cost effectiveness model. Specifically, the goals were to identify input variables (costs and effects) for estimating the CER and to identify/design valid, reliable, and sensitive instruments to measure these inputs. What constitutes costs and effects of youth EFNEP and how they should be measured were discussed in the workshop. Arguments for measurement of specific constructs were justified and passed by the panel members.

3.2.2.2. Decisions on Measuring Costs and Effects: The following decisions were made to precede the measurement of costs and effects.

- In order to ensure the comparability of the CER across common studies, the Panel on Cost effectiveness in Health and Medicine (PCEHM) recommends a methodological guide to conduct CEA (Gold et al. 1996). This guide was followed in this study in measuring both costs and effects as far as applicable.
- CEA are done with a perspective in mind, which is essentially determined by the purpose of the analysis. Because the key target audiences for the cost effectiveness model are EFNEP administrators, this study utilizes the program manager's perspective to CEA. CEA done from this perspective, although narrow, is appropriate in the present study where the program is being evaluated for its effectiveness in achieving its stated goals.
- Because lack of a universal instrument to measure the effects of youth EFNEP is one of the primary reasons for lack of evaluation studies on youth EFNEP, this project developed its own instrument to measure the effects of the youth EFNEP for CEA.

Details on conceptualization of cost effectiveness model for Virginia, and the decisions made in the workshop can be found in Serrano et al. 2011.

Time Frame for the Cost Effectiveness Analysis for Virginia

The timeframe for the cost effectiveness analysis was set to be one school calendar year, which ranged from September 2011 to May 2012. Costs data for the same period were collected from 15 EFNEP unit offices under which the program was implemented, and also from the State Office. Effects data were collected from the pre-post survey administered to all youth EFNEP participants in 15 counties of Virginia, during the same time period. The instruments used to collect both costs and effects data are given in the chapter appendix.

3.2.3. Measuring Effects – the Denominator

The measures of program effects, the denominator in equation (1), used in CEA are usually guided by the purpose of *the program. Given the purpose of the youth EFNEP program is to improve participants' practice to healthy nutritional behavior,* change in behavior was identified as one of the primary measure of program effects. Other potential outcome measures such as self-efficacy and behavior intentions were also explored.

Different models of individual health behavior suggest that behavior change is mediated by variables such as self-efficacy and behavior intention (Glanz et al. 2008). Self-efficacy (SE) is the person's confidence in performing a particular behavior (Glanz et al. 2008; Pajeres 2010). It is a belief one has that leads to a particular course of action. Self-efficacy affects the amount of effort put by an individual in taking that particular action and thus the behavior. Behavior intention, on the other hand, measures a person's strength of intention to perform a behavior and is determined by one's attitude towards performing the behavior and his subjective norm

associated with the behavior (Glanz et al. 2008; Ajzen 1991). The review of literature related to youth nutrition suggests that self-efficacy and behavior intention towards healthy nutrition are the most commonly used predictors of nutrition behavior (Vries et al. 1988; Vries et al. 1995; Parcel et al. 1995; Reynolds et al. 1993; Vereecken et al. 2005; Lein et al. 2002; Baranowski et al. 2000; Contento 1995). In the context of CEA and the goals of the program, these are clearly intermediate outcomes.

Based on the empirical evidence and suggestions from the advisory board, self-efficacy towards healthy nutrition, behavior intention towards healthy nutrition, and nutrition behavior were selected as three indicators of youth EFNEP outcomes/effects. These outcome indicators are latent constructs, which cannot be measured directly. In such a situation, several items to measure the different dimensions of the construct are formulated. The responses to each item serve as indicators of latent constructs (Emberston 2010; Wilson 2010).

Given the above background, the first step of the analysis was to develop a new standardized instrument for the purpose of measuring these outcomes of youth EFNEP. Details on the process of instrument development to measure program effects are discussed next.

3.2.3.1 (Step 1): Development of Instrument to Measure Youth EFNEP Effects

3.2.3.1. a. Literature Review of Evaluation Instruments

The process of instrument development began with an extensive review of literature to identify valid, reliable and sensitive instruments on diet and physical activity for youth, focusing on instruments developed specifically for limited resource youth and Extension-delivered programs. A total of 15 instruments were reviewed, of which 10 were specific to the youth

EFNEP.[2] Most of these instruments designed specifically for youth EFNEP evaluation were not tested (rigorously) for reliability and validity. The instruments focused only on some aspects related to nutrition. More general type of instrument which measures multiple aspects of nutrition which would be particularly beneficial and practical for a broader set of programs such as EFNEP was lacking (Hermandez-Garbanzo et al. 2013).

As a starting point, a pool of potential questions was developed based on the 15 selected existing instruments and was shared with the advisory board. Because, none of these instruments were specifically designed to capture the goals of youth EFNEP, the advisory board requested that the team develop a list of 'indicators' based on the Dietary Guidelines for Americans, to ensure that the instrument was comprehensive and 'could' possibly be used by other states.

3.2.3.1. b. Identification of Topics and Measurable Outcomes from the Dietary Guidelines

The Dietary Guidelines for Americans 2005 were reviewed and the entire lists of topics most appropriate for low-income youth ages 7 to 14 were selected. Out of over 30 guidelines addressing topics such as alcoholic beverages and sodium intake, 15 were selected. In addition, topic areas that the dietary guideline specifically encompassed were identified. The topics were presented to the advisory board who then voted the selected topics in terms of relevance priority, and its appropriateness for the population under study. In addition, advisory board members were asked to identify topics that they felt were implicit in the guideline and thus should be measured. The following topics emerged from the discussion:

- different food groups from MyPyramid,
- sugar-sweetened beverages,

[2] A separate review article summarizing these instruments have been published (Hernandez-Garbanzo et al. 2013).

- discretionary calories,
- physical activity,
- portion size,
- sodium intake.

The advisory board members were then asked to align the selected topics with the outcome measures such as knowledge, skills, attitudes, behavior, to understand how the topics would be best measured. The advisory board members later agreed that instead of knowledge, skills, attitudes, and behaviors, the evaluation instrument should measure behavior, self-efficacy, and behavior intentions. The choice of measure followed because the overall goal of nutrition education programs is to impact behavior, and prior research had established that self-efficacy and behavior intentions are two mediators that are strongly linked to behavior outcomes among youths (Cerin et al. 2009). Therefore, the advisory board decided that the selected dietary guidelines should be measured in terms of behavior, self-efficacy, or behavior intention outcomes.

3.2.3.1. c. Construction of the Evaluation Instrument

New questions were generated as well as the questions identified from the literature review were modified to address all of the selected topics from the US Dietary Guidelines and framed in terms of self-efficacy, behavior intention, and behavior. Specifically, 111 questions were derived, covering the various topics. The advisory board was again asked to assess the questions. Given their feedback, questions with vague or inappropriate topics were removed from the list of questions. Definitions and clarifications were added when needed (i.e., definition for low-fat, etc.).

Lastly, the format of the response choice had to be determined. Two formats for asking questions and framing the response choices were common in the literature: i) “amount of food” and ii) “frequency of eating”. For example, the “amount of food” type response asks students how many ounces of whole grain you ate yesterday. On the other hand, the “frequency of eating” type question asks students, for example, how many times you ate cereals yesterday. Given the age of the target population (young children), tradeoffs between precision and accuracy of measurement by using both amount type and frequency type questions were discussed. The advisory board later decided to get feedback from Virginia youth EFNEP Program Assistants (PAs), to confirm the appropriateness of the proposed indicators and provide guidance on the response categories. Four highly effective youth EFNEP PAs from Virginia were consulted about whether the questions should be framed in terms of asking about the “amount of food” or the “frequency of eating” the food. All four PAs agreed that the response should be measured by frequency of food (e.g. times per week) rather than the quantity of food (i.e. cup, ounces), given the level of comprehension of the target group (youths). The PAs shared their experience that very few children sit down and actually eat a given amount of food at one time. In addition asking for the quantity (amount of food) would require kids to do something abstract and add up amounts over the course of a day, and their estimates would likely be imprecise. Therefore, the PAs recommended using frequency of eating type response format over the amount of food type response format. The PAs also helped eliminate as well as add some questions as based on feedback.

3.2.3.2 (Step 2): Pilot-Testing of the Evaluation Instrument

A pool of 93 items[3] was initially developed to measure self-efficacy, behavior intention, and behavior, each measured using 31 items. These items formed three separate sets of preliminary instruments. Similarly to other studies measuring self-efficacy, behavior intentions, and behavior, the items in all three instruments were the same except that the wording in the beginning of items was specific to the corresponding outcome.

In consultation with advisory board and program assistants in Virginia, the team recognized the response burden of 93 questions was too great for the targeted age group. Consequently the decision was made to pilot test all questions and to determine which were redundant and could be culled to a more feasible number.

Six youth EFNEP PAs pilot-tested the preliminary instrument with youth EFNEP participants in Virginia, during December 2010 to February 2011. Two sets of instruments were pilot -tested on each group of the participants over the course of two days. In the first day, all the participants were given the behavior component. On the second day, about half of the participants were given the self-efficacy survey and the other half of the participants received the behavior intention survey. A total of 272 surveys on behavior, 115 surveys on behavior intention, and 139 surveys on self-efficacy were completed.

3.2.3.3 (Step 3): Statistical Procedures for Culling Questions, and Reliability and Validity of Instrument

Data gathered from the pilot test was analyzed to select the best items from the preliminary instruments that would be used to develop the final instrument to measure the effects

[3] The pool of items covered both the positive and negative dimensions of self-efficacy, behavior intention and behavior. Negative dimension refer to questions about reducing the intake of unhealthy food, and positive dimension refer to questions about increasing the intake of healthy food. Conventionally, only positive items were used to measure self-efficacy and behavior intention.

of the program. A conventional method based on classical test theory; item analysis, was conducted to select best items from the pool of items. Selection of items was later guided by another statistical method called exploratory factor analysis.

Item Analysis: Item analysis is a commonly used method in psychometrics for developing instruments. It plays an important role in construct validation. Item analysis extracts items that measure a uni-dimensional construct by exploring how the test takers respond to each item, and how the items relate to their overall performance (Nunnally and Bernestein 1994). The decision about good and bad items is based on the item-total correlation statistics between the scores of an individual item and the total score on the test excluding that item. This measure checks whether a given item is consistent with the other items in the instrument. If a given item is consistent with the overall test, the item-total correlation should be high. A small item-total correlation means that the item is not measuring the same underlying construct as the rest of the items, thus should be revised or discarded. Item analysis can be used to determine the most valid questions for the development of a shorter, efficient questionnaire.

Using the software "jMetrik", separate analyses were done for the three sets of preliminary instruments: self-efficacy, behavior intention, and behavior. We started with all 31 items in the first iteration. As a rule of thumb, any items that do not correlate strongly ($r < .30$) with the total test score were regarded as bad items (Field 2005). Items with a negative or low correlation in each iteration were discarded until all the remaining items had an item total correlation > 0.3.

Among 31 items, 11 met the criteria of good items in each instrument. Results of item analysis are given in Appendix 2.2 (Tables 12, 13, and 14). Good items that measured self-

efficacy and behavior intention were all negative items, and the good items that measured behavior were all positive items. The findings suggested that the preliminary instrument was probably measuring multiple constructs. Violation of uni-dimensionality would have been the cause for culling entire positive items in the self-efficacy and behavior intention instrument, and all negative items in the behavior instrument. Rather than do this, exploratory factor analysis (EFA) was used to explore the number of underlying constructs (factors) in the preliminary instrument.

Exploratory Factor Analysis: In EFA, all the items are modeled as linear combinations of potential factors that have common characteristics. The coefficient of each of the items, known as factor loadings, gives the relative contribution of an item to make a factor. Items with higher factor loadings are important variables for explaining a particular factor. If all the items have high correlations with only one factor, this provides evidence that the instrument is measuring a uni-dimensional construct. EFA not only identifies the number of latent factors but also identifies good items that measure the underlying construct. The purpose of EFA was thus twofold: first, to check dimensionality (that is whether the instrument is measuring one or more underlying latent constructs), and second, to select the best items to measure the underlying construct. Statistical software SPSS was used to perform EFA.

EFA for checking dimensionality: There are different, often debated, criteria in deciding the number of factors being measured by a set of instrument. According to Kaiser's criterion, any component with eigenvalue >1 is considered as a separate factor. The eigenvalues when graphed against the factors with which it is associated, a scree plot, shows the relative importance of each factor. The scree plot provides useful information regarding the number of important factors

being measured. The cutoff point for selecting the number of important factors is at the point of inflection of the scree plot (Field 2005). The scree plots for all three preliminary set of items are given in Appendix 2.3 (Figures 4 and 5). Clearly, scree plots derived using pilot test data show that there are two main underlying factors for each of the preliminary instruments, which provides evidence that each instrument measured more than one constructs. To further explore the multi-dimensionality of the instrument, and for item selection, exploratory factor analysis was conducted.

EFA for item selection: The relationship between the items and the factors, are measured as factor loadings. The factor loadings can be thought of as correlation coefficients or sometimes as regression coefficients between a factor and the items. The interpretability of factors can be enhanced by the technique called rotation (Field 2005). An oblique rotation was used which allows the underlying factors to be correlated. The higher the factor loading, the more variance in the factor is explained by the items. In selecting the good items, items with factor loadings > 0.3 were sorted as it is typical to consider factor loadings with this threshold value (> 0.3) to be important (Field 2004). Results from EFA are presented in Appendix 2.3 (Table 15)

Exploratory factor analysis extracted two main factors for each outcome: self-efficacy, behavior intention, and behavior. Most of the positive items appeared to be constituents of one factor, while most of the negative items of another factor. A few items did not correlate (had r < 0.3) with any of the two factors. This result was consistent with all three sets of preliminary instrument.

Based on these results, selection of good items based on the item analysis may not be a sound method for measuring multi-dimensional latent constructs. Therefore, results from the

factor analysis were used to select items for the final instrument based on their factor loadings. All items with factor loadings >0.3 were regarded as 'good' and used in the final instrument. Unfortunately, the resulting instrument however would still contain more than 50 items. It was recognized that the high number of items would substantially increase the response burden to the target group. After another review of the literature and a prolong discussion with "advisory board" members, the following decisions were made that finalized the instrument.

- Because a review of instruments that measure the effects of youth nutrition education program in the literature revealed that self-efficacy is more widely used than behavior intention, the behavior intention indicators were discarded from the final instrument.
- Most studies published in nutrition education focus on positive dimensions of self-efficacy. Therefore, negative items were discarded.

The final instrument included only two indicators: self-efficacy, and behavior. The numbers of total items were reduced to 35 items: 19 items to measure behavior, and 16 items to measure self-efficacy. In addition, questions regarding participants' demographic information, and food availability at home were collected. The final instrument is attached in Appendix 3.

Measuring Total Effects of the Program

A common and practical method to measure the effects of an educational program is to use a *pre* and *post*-test approach. The effects instruments were administered *pre* and *post* intervention to measure changes in responses for each participant. Improvements in individual participant's responses have to be amassed to generate the total effects for the program which is required to calculate the CER (denominator of equation 1). Two alternative approaches to generate total effects are possible.

- **Score based measures**: This measure uses the raw scores of individuals on all items and measure the change in raw scores from pre to post test. Standardized tests, such as GRE and SAT, are common examples of score based measures, which have been used as a measure of educational achievement (for example, Congressial Budget Office (CBO), 1987).
- **Count based measures**: This measure counts the number of individuals satisfying some criteria. These are very common in the education field. The "No Child Left Behind" uses the percentage of students achieving proficiency in a given grade and subject; National Assessment of Education Progress uses the percentage of students meeting or exceeding the national standards, both of which are basically the count based measures (Virginia Department of Education, 2012; US Department of Education National Assessment of Title I, final report, 2007).

Simple algebra reveals that improvements in total scores does not map one-to-one with changes in numbers of individuals improving in outcomes. In fact, it is possible for the total (or average scores) to improve while the percentages of individuals improving actually decreases. Because the objective of the program is to improve numbers of individuals (counts) not scores, the count based measure was favored over the score based measure and this decision approved by the advisory board. The count based measure is used in other areas, such as education and poverty (for example, Heck 2006) and is standard in the poverty literature (for example, Jolliffe et al. 2005). It is also consistent with the current practice of the USDA in measuring EFNEP outcomes which counts the number of individuals who improve on *at least one item* among a list of items. Effectively, in this chapter, the denominator of the CER is measured by two measures: 1) the total numbers of participants whose response in at least one of the 19 items measuring

behavior improved, and 2) the total numbers of participants whose response in at least one of the 16 items measuring self-efficacy improved.

3.2.4. Measuring Costs – the Numerator

The underlying principle in measuring costs of the youth EFNEP program is that the estimates of costs should reflect the value of resources that are used in the provision of the program (HWHK). Only the direct costs associated with implementing the program are considered. Indirect costs borne by program participants are ignored because the cost effectiveness analysis is done from program managers' perspective. The direct costs of implementing the HWHK program are divided into four main categories: 1) Labor Cost; 2) Capital Cost; 3) Material Cost; 4) Energy Cost.

Each cost category is a composite of multiple components. The cost attributed to the HWHK should be reflective of the share of the HWHK program within the EFNEP and the Virginia Cooperative Extension (VCE) under which the program are usually implemented. Thus, adjustments are made where appropriate. Further, the costs are prorated for months in which the program is actively implemented within the time frame for which the CEA is done.

3.2.4.1. Labor Cost

The cost related to the human resources directly involved in delivering and managing the youth EFNEP are considered as labor cost. Youth EFNEP program assistants (PAs) are the key personnel involved in the management and the implementation of the HWHK program. The EFNEP administrators and program support staffs are also involved in managing the HWHK program, but their contribution to the HWHK is relatively small. Therefore, labor costs

associated only with the youth EFNEP PAs are considered. Labor costs include expenditures on salaries and benefits, travel, and training.

3.2.4.1. a. Salaries and Benefits

Salaries and benefits consist of the compensation provided to the PAs in the form of salaries and fringe benefits. Some youth PAs are responsible for the youth EFNEP only, while others are responsible for both the youth and the adult EFNEP program. Consequently, the share of time allocated to adult programs is not included. Furthermore, within the HWHK, PAs spend their time in various sub-activities such as recruiting youths, studying and preparing for the classes, preparing food for demonstration, teaching/delivering the HWHK curriculum, traveling, and filling out forms for administrative purposes. Within the youth EFNEP, the youth PAs deliver other curricula, for example, Literacy, Education, Activity Program (LEAP), Organ Wise Guys, Teen Cuisine, and Professor Popcorn, in addition to the HWHK curriculum in Virginia.

The share of salaries and benefits attributed to the HWHK program are reflective of the share of time spent by each PA on the provision of the HWHK curriculum only. Further, salaries and benefits are only measured for months in which the actual HWHK program is implemented, that is, only for months active in the HWHK. Thus, the annual salaries and benefits of the youth PAs are adjusted for the share of their work time spent on the HWHK, and for months active in the HWHK program. The total labor cost for salaries and benefits attributed to the HWHK are derived as follows:

Cost of salaries and benefits[4] = *Annual salaries and benefits of the youth PAs * Share of PA work time spent on the HWHK * Proportion of months active in the HWHK program*

3.2.4.1. b. Training

Training costs include costs associated with the training of the youth EFNEP PAs. It consist of cost associated with training PAs after their initial recruitment into the youth program, and cost associated with attending all trainings/workshops, related to the HWHK program, attended by the youth PAs within the time frame in which the program evaluation is done.

Cost of training[5] = *Costs associated with youth PAs initial training after recruitment + Costs associated with attending trainings/workshop/meetings during the time frame of program evaluation*

3.2.4.1. c. Travel

Travel costs include costs associated with PAs travel in order to manage and implement the HWHK program. Some of the common causes for travel include school visit for student recruitment, purchase of supplies, curriculum delivery, attending meetings and workshop. The product of the numbers of miles traveled by PAs in their personal vehicles in order to deliver the program and the rate of reimbursement for travel yields the travel cost incurred for each PA.

Cost of travel[6] = *Numbers of miles traveled by PAs in their personal vehicles in order to deliver the HWHK program * Rate of reimbursement for travel cost per mile*

3.2.4.2. Capital Cost

[4] Source: State EFNEP office and PAs bimonthly time allocation survey
[5] Source: State EFNEP office
[6] Source: State EFNEP office

Capital costs are costs associated with the capital resources directly used in support of delivering and managing the HWHK program. Only the costs of capital resources used at the county level are considered. This is because the share of capital attributed to the HWHK is relatively small at the EFNEP offices higher in the hierarchy (such as district level and state level). Office space and equipment (for example, computers and copiers) are the main capital resources used for the HWHK.

3.2.4.2. a. Office Space

Office space includes the cost of office space used by the staffs directly involved in the HWHK program. The HWHK program staffs at the county level are housed at the EFNEP unit offices. Most of the EFNEP unit offices are usually housed within the Cooperative Extension offices. The Cooperative Extension office building is often shared by multiple agencies. Most of the Cooperative Extension office buildings are owned by the local government and some of them are rented. The measurements of cost of office space for the office housed in the owned and in the rented buildings are different, hence are dealt separately.

3.2.4.2. a. 1 Owned Office Space

The cost of owned office space is derived from the current market value of the building in which the program is housed. The Office of Management and Budgeting (OMB) Circular 87 on "Cost Principles for State, Local and Indian Tribal Governments" states that for buildings owned by local government agencies, the cost of space should be computed as an annual rate not exceeding 2% of total costs of acquisition. Following the guideline, only 2% of the current value of the building is used as the cost of office space for a given year.

The cost of office space attributed to the youth EFNEP program should be reflective of the office space actually used by the youth program only. Further, the value of office space should only be considered for months active in HWHK. Thus, the market value of the office building in which the program is housed is adjusted by the share of office space used by the program, and for months active in the program.

Cost of Owned Office Space[7] *= Current market value of building where EFNEP office is housed *annual usage fee as established by the OMB * the percentage of building space occupied by EFNEP * percentage of EFNEP space occupied by youth EFNEP *Proportion of months active in the HWHK program*

3.2.4.2. a. 2 Rented Office Space

The cost of rented office space for the program is derived from the value of annual rent paid for the building in which the program is housed. Similar to the owned office space, the cost of office space attributed to the program should reflect the office space used by the youth EFNEP program. Also, the cost of office space should be reflective of months active in the HWHK program. The value of monthly rent paid for the entire office building in which the program is housed, for the months active in the HWHK, is adjusted by the share of office space used by the program.

Cost of Rented Office Space[8] *= Annual rent of the building where EFNEP is housed * the percentage of building space occupied by EFNEP * percentage of EFNEP space occupied by youth EFNEP * Proportion of months active in the HWHK program*

[7] Source: Unit offices (cost survey done with Unit Office Incharge); and PAs bimonthly time allocation survey
[8] Source: Unit offices (cost survey done with Unit Office Incharge); and PAs bimonthly time allocation survey

3.2.4.2. b. Equipment

Equipment cost includes the value of equipment such as computers, printers, fax machines, copiers, VCRs, cameras used by the youth EFNEP staffs for the HWHK program. Only the values of equipment at the unit offices are considered. The OMB guideline (Circular A 87) allows 6.67% of the total cost of equipment to be used each year. Following the guideline, only 6.67 % of the cost of equipment is used as annual cost of equipment.

Some of the equipment is solely used by the youth EFNEP, whereas other equipment is shared by the adult and the youth program. For equipment that is not solely used by the youth EFNEP, the cost attributed to the program is adjusted by the share of equipment used for the youth EFNEP. Further, the cost of equipment is adjusted for months active in the HWHK.

Cost of Equipment[9] = *Sum over all equipment type (total cost of equipment * Share of equipment used for the youth EFNEP *Annual usage fee as established by the OMB* Proportion of months active in HWHK)*

3.2.4.3. Material Cost

Material costs are costs associated with any material resources used in support of delivering the HWHK program. Two types of materials used for the HWHK are identified: supplies (for example, food) and printed materials.

3.2.4.3. a. Supplies

[9] Source: Unit offices (cost survey done with Unit Office Incharge); and PAs bimonthly time allocation survey

The cost of supplies includes the cost of food items purchased by the PAs for demonstration use in support for the delivery of the HWHK curriculum. PA's are reimbursed for the expenses incurred in purchasing the supplies for the program.

Cost of Supplies [10] = *Cost of supplies purchased by PA for the demonstration purpose for the HWHK for months active in HWHK.*

3.2.4.3. b. Printing

Printing cost includes the cost of printing the evaluation instruments, handouts and any other materials used in support of implementing the HWHK program. The number of students who receive each type of printed materials multiplied by the unit cost of printing those materials yields the cost of printing.

Cost of Printing[11] = *Number of youth EFNEP participants * unit cost of printing*

3.2.4.4. Energy Cost

Energy costs are costs associated with the usage of energy in the EFNEP unit offices, in support of the HWHK program. It mainly consists of the cost of utilities.

3.2.4.4. a. Utilities

The cost of utilities includes the value of money spent on utilities such as phone, electricity, water, sewage, garbage, heating oil/gas and janitorial or maintenance services. The cost of utilities is derived from the monthly bills paid for each utility type. Utility bills are usually available for the entire building in which the youth EFNEP is housed. Energy costs

[10] Source: State Office (As reported in the "green system")
[11] Source: State Office

attributed to the program should reflect the share of utilities consumed by the youth EFNEP program. Because the share of energy use is difficult to measure, the share of office space occupied by the youth EFNEP program is used as a proxy for the share of utility usage. The costs of utilities are also adjusted for months active in the HWHK program. The adjusted monthly utility bills for each utility type are summed over months active in the HWHK to get the cost of utilities.

Cost of Utilities[12] = *Sum of the monthly cost of utilities of each type, for months active in HWHK, for the entire building in which youth EFNEP is housed * percentage of building space occupied by EFNEP* percentage of EFNEP space occupied by youth EFNEP*

Measuring Total Costs of the Program

Costs under each category are measured separately for individual unit offices (county offices from where the program is delivered) and added to get the total cost of program for each county. Total cost for each county is then summed across all unit offices to get the total cost of the youth program in Virginia.

Total Cost of the Program = *Sum over all counties (Labor Cost + Capital Cost + Material Cost + Energy Cost)*

3.3. Results and Discussion

To demonstrate the application of CEA tools and methods developed in section 3.2, data from youth program in Virginia are used. The results discussed below are derived from the costs and effects data collected from the following 15 counties in Virginia: Appomattox, Carroll,

[12] Source: Unit offices (cost survey done with Unit Office In-charge); and PAs bimonthly time allocation survey

Charles City, Culpeper, Hampton, King William, Loudoun, Lynchburg, Norfolk City, Patrick, Richmond City, Rockingham, Scott, Westmoreland, and Wise.

3.3.1. Costs of Implementing Youth EFNEP- HWHK in Virginia

The total costs of implementing the youth EFNEP- HWHK program in 15 counties of Virginia during the school calendar year 2011/2012 was estimated to be $134,333 (Labor = $112,558; Capital = $7,250; Material = $5163; Energy = $9362). Cost of labor (salaries and benefits including travel and training) had the largest share of expenditure for the program (84%), followed by energy cost (7%), capital cost (5.4%), and material cost (3.8%). The costs of implementing the program in each of the 15 counties are summarized in Table 1. The average total cost of implementing a youth EFNEP program in one county is $8,956. The range of total costs across programs varied from $1,982 to $16,825. Similarly, the average cost related to labor was $7,504 (range $1,806 - $16,673); the average cost related to capital was $604 (range $30 - $2,283); the average cost related to material was $344 (range $27 - $1,651); the average cost related to energy was $780 (range $4 - $2,119).

[Place Table 1 Approximately Here]

Cost data for some of the cost headings were not available for some counties, therefore not included in total cost calculation. Richmond and Carroll counties did not provide cost data for office space, equipment and utilities. Hampton County did not provide cost data on equipment. The youth EFNEP office at Norfolk is housed in some other governmental building, and the EFNEP is exempt from paying money for office space and utilities. So, youth EFNEP program in Norfolk incurs zero cost related to office space and utilities (rather than missing information). Loudoun County reported that the equipment used for the youth EFNEP program is

provided by the State and that the program does not spend on equipment. Missing data on cost is expected not to significantly change the total cost estimates because the cost share of capital cost and utility cost is relatively small for all programs in general. The impact of missing cost data in part will be handled in the sensitivity analysis.

3.3.2. Effects of Youth EFNEP-HWHK in Virginia

The effects of youth EFNEP-HWHK program were measured by two outcome indicators: change in nutrition related behavior, and change in nutrition related self-efficacy. Changes in behavior were measured by the total number of individuals who improve in at least one of 19 behavior items, from pretest to posttest. Similarly, changes in self-efficacy were measured by the total number of individuals who improve in at least one of 16 self-efficacy items, from pretest to posttest. These measures of effects are consistent with the current practice of the USDA in measuring total effect of the EFNEP.

A total of 2,566 students were reached by the program. Of students who were reached by the program, 73% of them (1,864) were evaluated. The rest of the students were excluded from the study because either they did not complete one or both of the pre/post surveys, or their pre/post surveys did not match based on the identifiers used.

One of the key assumptions while using change in behavior and change in self-efficacy as measures of effects is that no factors other than the program itself affects change in participants' behavior or self-efficacy. Given that the design of the program is not experimental, the influence of factors other than the youth EFNEP program on the outcome measures cannot be ruled out. In an attempt to control for some of these factors, the probability of a change in outcome (probability of participants improving in at least one item) was modeled as a function of

participant characteristics (gender, age, ethnicity, family type, geographic location), and program characteristics (class size, class type, number of volunteers assisting the class, duration of program, number of handouts distributed, PA's age and experience) (Table 2). Predicted probabilities for each participant only associated with the program characteristics were then generated by setting the coefficients on variables related to student characteristics equal to zero, which is equivalent to predicting the probability just associated with the program effects. The sum of these predicted probabilities across all individuals would then give the predicted number of people who improved controlling for the other confounding effects, at least those included in the model and these are called the predicted effects. Results using the predicted effects are compared with the results using the observed effects (actual improvements in behavior).

[Place Table 2 Approximately Here]

Change in Behavior and Self-efficacy:

The effects for each of the 15 counties are summarized in Table 3. On average, 171 participants were reached by the HWHK program in each county (range 26 - 434) of which about 124 students were evaluated (range 11 – 296). The average observed improvements in behavior was 119 (range 11 – 29), and the average observed improvements in self-efficacy was 119 (range 10 – 285). The predicted improvements were slightly higher. Of 1,864 students evaluated, 1,786 students (96%) improved in at least one behavior (observed effects) from pretest to posttest. About 1,782 students (96%) improved in at least one self-efficacy (observed effects) items from pretest to posttest. Accounting for the factors outside of program's control, the predicted effects are higher indicating that confounding factors may be offsetting program effects for some

individuals. For behavior the predicted effects (improvements) are 1819 and for self-efficacy 1853.

[Place Table 3 Approximately Here]

3.3.3. The Cost Effectiveness Ratio

Using the data on costs and effects collected from 15 counties, the CERs for youth EFNEP-HWHK in Virginia were computed and are reported in Table 4. The total CER (TCER) per participant reached by the program in Virginia was $52. For both outcomes, improvements in behavior and self-efficacy, the TCER per outcomes (actual effects) were on average $75. Using the predicted effects, the TCER per behavior improvement was estimated to be $75, and the TCER per self-efficacy improvement was $73.

The average CER (ACER) per participant reached is $89 (range $20 – $243). The ACER per improvement in behavior is $149 (range $24 – $537), and the ACER per improvement in self-efficacy is $153 (range $25 – $591). The values of ACER estimated by using the predicted improvements are slightly less. Estimates of ACER are higher than the estimates of TCER. The condition in which ACER is equal to TCER is only when the effects are equal for all counties (conditions derived in section 3.2), which is not the case here. The difference between the estimates of ACER and TCER basically reflects the variability in effects across counties.

[Place Table 4 Approximately Here]

3.3.4. Sensitivity Analysis

The estimates of the CER presented above are the point estimates and may suffer uncertainties. The uncertainties in the estimates of costs and effects could arise due to missing data for some cost components, and also due to possible measurement errors, which must be

accounted for. It is a common practice to exhibit the uncertainty of point estimates by constructing confidence intervals around the estimates. Confidence interval captures the variability in the estimate, the likelihood that the sample drawn from the same population will yield the estimated value. Bootstrapping is a common method of choice to determine the variability of a ratio estimator (Chaudhary and Stearns 1996; Briggs et al. 1999; Polsky et al. 1997). The standard errors and the confidence intervals of the estimated TCER were derived by using a non-parametric bootstrapping method. First a sample of 15 sets of costs and effects were drawn from all observations of costs and effects using a simple random sampling with replacement. The bootstrap replicate of the TCER was then computed from the drawn sample. The method was repeated for 10,000 times to obtain 10,000 independent replications of bootstrap TCER. The observed distribution of the standard error and the confidence interval of bootstrap TCER are presented in Table 5. The sampling distribution of the TCER estimates from each bootstrap replications are plotted in Figure 1.

The estimated range of 95% confidence interval is $56 to $115 and $55 to $114 for TCER per improvement in observed and predicted behavior, respectively. Similarly, the 95% confidence interval for TCER per improvement in observed and predicted self-efficacy is $56 to $116 and $53 to $111, respectively (Table 5). The observed values of TCER for both outcome measures lie within the estimated confidence range of TCER.

[Place Table 5 Approximately Here]

[Place Figure 1 Approximately Here]

3.3.5. Cross-County Comparison of CER

The point estimates of the CER by county serve as a starting point for making a comparative analysis in identifying the most cost effective program, and are given in Figure 2. Three measures of effects for each county are shown. The first bar (blue bar) measures the cost per student reached, which is simply the average cost per participant reached by the program. Second and third bars give the cost per improvement for behavior and self-efficacy, respectively. For each county, the numerators for all three outcomes are same and what varies among the three is the measure of effects.

[Place Figure 2 Approximately Here]

The graph clearly shows a huge variation in the estimates of CER across counties. Counties with high CER per reach indicate that such counties (counties 5, 9, and 15) spend more dollars per participant. This could be because such counties either have high program cost or they recruit only a small number of participants. The within county gap in the estimates of CER per reach and CER per improvements is staggering for some counties, and this is an indication of program ineffectiveness. On average, the CER per improvement is about 1.7 times the CER per reach. For county 2 the CER per improvement is only 1.2 times the CER per reach. On the other hand, for county 9, the CER per improvement is about 2.4 times the CER per reach. For counties 1-4, 6-8, and 11 the cost per reach is not very different than cost per improvement. For counties 9 and 5 the cost per improvement is much higher than cost per reach. Large discrepancy between CER per reach and CER per improvements means that relatively few program participants actually improve their behavior. The smaller the discrepancy, the more effective the program is in achieving the goal of improving the behavior of its participants.

So far, in estimating the CER, what counted as an improvement for an individual was that they improved on *at least one item*. The USDA improvement threshold for the adult EFNEP using the 10 item behavioral checklist is 1, and is followed in the current study to be consistent with the USDA measure. One can argue that the improvement measure used by the USDA (and consequently this study) has a very low bar in improvement threshold, and it completely ignores the degree of improvements achieved by the program. The improvement thresholds can be set higher by redefining *the number of items an individual improves on*. The improvement threshold could be any number between 0 and the total number of questions in the survey. For example, there are 19 items that measure behavior, and one could calculate the number of people who improved on 0 questions, 1 question, 2 questions, up to all 19 questions. Increasing the *improvement threshold* has implications on the estimates of CER. The total CER estimated by setting different thresholds on improvements are given in Table 6 (and Figure 3).

[Place Table 6 Approximately Here]

[Place Figure 3 Approximately Here]

The numerator of CER (total costs), used in estimating CER in Table 6, remains the same for all threshold values. Only the denominator (total effects) is changed by increasing the improvement thresholds. Obviously the CER estimate increases with higher threshold which sets more stringent criteria to be counted as improvements. The estimates of CER shows that rate of increase in CER is not linearly related to the increase in threshold criteria. The CER increased from $75 per improvement to $78 per improvement when threshold increased from 1 to 2, whereas the CER increased from $209 to $306 when threshold increased from 7 to 8. The increase in CER is small at smaller values of thresholds, however, it takes offs rapidly past

threshold 5. For higher threshold values, the increase in CER is much higher for self-efficacy improvements compared to behavior improvements.

In a nutshell, we can conclude that setting more stringent thresholds on improvements rapidly increases the CERs, and is more so for higher thresholds. This finding has important policy implications for improving program effectiveness. By increasing threshold, only the individuals who improve in multiple dimensions related to nutrition are counted as program effects (improvements). Also, the improvements in more than one dimension can be achieved with only a small increase in costs, for example, by increasing costs from $75 to $78, gains in outcomes or program effectiveness can be doubled (from 1 to 2 improvements). On the other extreme, if the program aims to improve the behavior in all 10 dimensions (threshold = 10), the program cost will increase to a very large extent ($736 per improvement). The problem the policymakers now face is the obvious tradeoffs between the cost and the desired program outcomes. If the policy objective is to secure improvements in multiple dimensions, then the program will certainly incur higher costs. Deliberation however is to be made regarding the choice of optimal threshold for improvements.

3.4. Conclusions

A model for evaluating the cost effectiveness analysis (CEA) for youth EFNEP program is developed in this chapter. The challenges due to lack of general, valid, and reliable tools for measuring outcomes of the youth EFNEP was surpassed with the help of the advisory board members. The instrument developed in this chapter is general enough to capture the various aspects of nutrition education and is based on the Dietary Guidelines for Americans. This makes

it readily usable for evaluating other youth EFNEP programs across the nation. The procedures and methods developed to measure both costs and effects of youth EFNEP in this chapter opens up avenues for other youth nutrition education programs in other states to conduct cost effectiveness analysis of their respective programs. The procedure for conducting CEA developed in this chapter is applied to Virginia youth EFNEP program for demonstration.

The cost effectiveness analysis of youth EFNEP program (HWHK curriculum) in Virginia was done for one school year (2011/12). From a program managers' perspective, data on cost related to the management and implementation of the program were collected from 15 cooperative extension unit offices under which youth EFNEP was implemented, in the given time frame. Data on effects were collected from all participants participating in the youth EFNEP program in the same timeframe from all 15 counties. Total costs and total effects of the program were compared. The total costs of the program were estimated to be $134,333, and the total effects of the program measured by the number of participants whose behavior improved was 1786 (predicted 1819), and that measured by the number of participants whose self-efficacy improved was 1782 (predicted 1853).

The total CER (TCER) for the youth EFNEP program in Virginia was estimated to be around $75 per improvement in behavior and about the same for improvement in self-efficacy. Now, how good the estimated figures of CER are is unknown because these are the first estimates of its kind for youth EFNEP. Cross county comparisons reveal a wide variation in the estimated CER across counties. The choices of thresholds on improvements are also shown to have important implications on the estimates of CER.

In a separate paper (Baral et al. 2013), we estimated the maximum average cost of improving *adult* EFNEP participant's behavior on three outcome indices: food resource management practices (FRMP), nutrition practices (NP), and food safety practices (FSP). We found that, for Virginia, the maximum average CER to be $579, $544, $718 per improvement in FRMP, NP, and FSP, respectively. While these were the maximum average cost estimates for adult program for Virginia, these are the best available information that could be used for comparison. It is safe to say that the youth EFNEP program in Virginia with much lower estimates of CER ($75 per improvement) are highly cost effective compared to the adult program in Virginia.

With a tightening federal budget, there is an increasing pressure to identify and use cost effective methods to achieve the stated goals of the youth EFNEP. There can be no discussion and comparison of cost effectiveness without estimates. This study developed the tools and procedures to estimate the cost effectiveness of youth EFNEP and estimated the CER for Virginia. The tool and procedures developed in this study can be utilized by other youth EFNEP programs to generate information on the cost effectiveness for respective programs. Results then can be used by the state and national EFNEP administrators to demonstrate program success, for comparing the cost effectiveness across states and ultimately to improve EFNEP efficiency.

References

Ajzen, I. 1991. Theory of Planned Behavior. *Organizational Behavior and Human Decision Process,* 50: 179-211.

Baranowski, T., M. Davis., K. Resnicow., J. Baranowski., C. Doyle., L.S. Lin., M. Smith., and D.T. Wang. 2000. Gimme 5 Fruit, Juice, and Vegetables for Fun and Health: Outcome Evaluation. *Health Education & Behavior*, 27 (1): 96-111.

Baral, R., G. C. Davis., S. Blake., W. You., and E. Serrano. 2013. Using National Data to Estimate Average Cost Effectiveness of EFNEP Outcomes by State/Territory. *Journal of Nutrition Education and Behavior*: 45(2): 183-187.

Briggs, A.H., C.Z. Mooney., and D.E.Wonderling. 1999. Constructing Confidence Intervals for Cost-Effectiveness Ratios: An Evaluation of Parametric and Non-Parametric Techniques Using Monte Carlo Simulation. *Statistics in Medicine*, 18: 3245-3262.

Cerin, E., A. Barnett., and T. Baranowski. 2009. Testing Theories of Dietary Behavior Changes in Youth Using Mediating Variable Model with Intervention Programs. *Journal of Nutrition Education and Behavior*, 41:309-318.

Chaudhary, M.A., and S.C. Stearns. 1996. Estimating Confidence Intervals for Cost-Effectiveness Ratios: An Example from a Randomized Trial. *Statistics in Medicine*, 15: 1447-1458.

Congress of the United States, Congressional Budget Office. 1987. Educational Achievement: Explanations and Implications of Recent Trends. Available at http://www.cbo.gov/sites/default/files/cbofiles/ftpdocs/62xx/doc6204/doc13b-entire.pdf Accessed September 25, 2012.

Contento, I., G.I. Balch., Y.L. Bronner., L.A.. Lytle., S.K. Maloney., C.M. Oslon., and S.S. Swadener. 1995. The Effectiveness of Nutrition Education and Implications for Nutrition

Education Policy, Programs, and Research: A Review of Research. *Journal of Nutrition Education*, 27:277-418.

Dollahite, J. S., D. Kenkel, and C.S. Thompson. 2008. An Economic Evaluation of the Expanded Food and Nutrition Education Program. *Journal of Nutrition Education and Behavior*, 40:134-43.

Embretson, S.E. 2010. Ed. Measuring Psychological Constructs: Advances in Model-Based Approaches. 1st edition. *American Psychological Association*. Washington, DC.

Field, A. 2005. Discovering Statistics Using SPSS, 2nd Ed. London: Sage Publications.

General Accounting Office (GAO).2004. Nutrition Education: USDA Provides Services through Multiple Programs, but Stronger Linkages Among Efforts are Needed. Washington, DC: United States General Accounting Office.

Glanz, K., B. K. Rimer., and K. Viswanath. 2008. Eds. Health Behavior and Health Education: Theory, Research, and Practice, 4th edition. San Francisco, CA: John Wiley and Sons.

Gold, M., J. Siegel, L. Russell, and M. Weinstein. 1996. Eds, Cost-effectiveness in health and medicine, ed. O.U. Press. New York.

Heck, R. H. 2006. Assessing School Achievement Progress: Comparing Alternative Approaches. *Educational Administrative Quarterly*.42: 667-699.

Hemandez-Garbanzo, Y., J. Brosh., E.L.Serrano., and R. Bhattarai. 2013. Psychosocial Measures Used to Assess the Effectiveness of School-based Nutrition Education Programs: Review and Analysis of Self-report Instruments for Children 8 to 12 Years Old. *Journal of Nutrition Education and Behavior*, (Article in Press).

Jolliffe, D., C. Gundersen., L. Tiehen., and J. Winicki. 2005. Food Stamp Benefits and Child Poverty. *American Journal of Agricultural Economics*, 87: 569-581.

Joy, A. B., V. Pradhan., and G. Goldman. 2006. Cost-Benefit Analysis Conducted For Nutrition Education in California" *California Agriculture*, 60:185-191.

Lien, N., L. A. Lytle., and K. A. Komro. 2002. Applying Theory of Planned Behavior to Fruit and Vegetable Consumption of Young Adolescents. *American Journal of Health Promotion*, 16(4):189-97.

Nunnally, JC., and I.H. Bernstein. 1994. Psychometric Theory, 3rd ed. New York: McGraw-Hill.

Office of Management and Budgeting. 1997. Cost Principles for State, Local and Indian Tribal Governments (Circular A 87). Washington, DC.

Pajeres, E.2010. Overview of Social Cognitive Theory and of Self-Efficacy. Available at: http://www.emory.edu/EDUCATION/mfp/eff.html . Accessed on 16 May 2010.

Parcel, G.S., E. Edmundson., C. L. Perry., H. A. Feldman., N. O'Hare-Tompkins., P.R. Nader., C. C. Johnson., and E.J.Stone.1995. Measurement of Self-Efficacy for Diet-Related Behaviors Among Elementary School Children. *Journal of School Health*, 65(1):3-40.

Polsky, D., H. A. Glick., R. Willke., and K. Schulman. 1997. Confidence Intervals for Cost-Effectiveness Ratios: A Comparison of Four Methods. *Health Economics*, 6: 243-252.

Rajgopal, R., R. H. Cox., M. Lambur., and E.C. Lewis. 2002. Cost-Benefit Analysis Indicates the Positive Economic Benefits of the Expanded Food and Nutrition Education Program Related to Chronic Disease Prevention. *Journal of Nutrition Education Behavior*, 34:26-37.

Reynolds, K.D., A.W. Hinton., R.M. Shewchuk., and C.A. Hickey. 1999. Social Cognitive Model of Fruit and Vegetable Consumption in Elementary School Children. *Journal of Nutrition Education,* 31(1): 23-30.

Schuster, E., Z. L Zimmerman., M. Engle., J. Smiley., E. Syversen., and J.Murray.2003.Investing in Oregon's Expanded Food and Nutrition Education Program (EFNEP): Documenting Costs and Benefits. *Journal of Nutrition Education Behavior,* 35: 200-6.

Serrano, E., S. Diehl., K, Jamison., J, Shelhamer., J, Wages., and K., Gehrt. 2007. Healthy Weights for Healthy Kids: How Virginia Cooperative Extension is Addressing Childhood Overweight, in Food, Nutrition, and Health. *Virginia Cooperative Extension,*.348-275

Serrano, E., M. McFerren., M. Lambur., M. Ellerbock., K. Hosig., N. Franz., M. Townsend., S. Baker., P. Muennig., and G. Davis. 2011.Cost-Effectiveness Model for Youth EFNEP Programs: What Do We Measure and How Do We Do It?" *Journal of Nutrition Education and Behavior,* 43 (4): 295-302.

Townsend, M. S., M. Johns., M. K. Shilts., and L. Farfan-Ramirez.2006. Evaluation of a USDA Nutrition Education Program for Low-Income Youth. *Journal of Nutrition Education and Behavior,* 38: 30-41.

United States Department of Agriculture (USDA). 2013.National Institute of Food and Agriculture. *Expanded Food and Nutrition Education Program*. Available at http://www.nifa.usda.gov/nea/food/efnep/efnep.html Accessed February 26, 2013.

USDA Cooperative State Research, E., and Extension Service. 2008. NC_TEMP211 Multi-State Team, EFNEP Related Research, Program Evaluation and Outreach: Statement of Issue(s) and Justification.

U.S. Department of Education. National Assessment if Title I Final Report. October 2007. Available at http://ies.ed.gov/ncee/pdf/20084012_rev.pdf Accessed September 25, 2012

Vries, H. D., E. Backbier., G. Kok., and M. Dijkstra. 1995. The Impact of Social Influences in the Context of Attitude, Self-Efficacy, Intention, and Previous Behavior as Predictors of Smoking Onset. *Journal of Applied Social Psychology*, 25(3): 237–257.

Vries, H. D., M. Dijkstra., and P. Kuhlman. 1988. Self-efficacy: the Third Factor Besides Attitude and Subjective Norm as a Predictor of Behavioral Intentions. *Health Education Research*, 3(3):273-282.

Vereecken, C.A., W.V. Damme., and L. Maes. 2005. Measuring Attitudes, Self-Efficacy, and Social and Environmental Influences on Fruit and Vegetable Consumption of 11- and 12-Year-Old Children: *Reliability and Validity*. 105(2):257-261.

Virginia Department of Education. 2012. Adequate Yearly Progress Reports. Available at http://www.doe.virginia.gov/statistics_reports/accreditation_ayp_reports/ayp/index.html Accessed September 25, 2012.

Wessman, C. B., and H. Jensen. 2002. An Evaluation of the Costs and Benefits of Iowa's Expanded Food and Nutrition Education Program. Final Report., in PM1865a, *I.S.U. Extension*, Editor.

Wilson, M. 2010. Constructing Measures: An Item Response Modeling Approach. New York, NY: Routledge.

Table 1: Cost of Youth EFNEP by County

County ID	Labor Cost		Capital Cost		Material Cost		Energy Cost		Total Cost
	Cost (in $)	Cost share	Cost (in $)	Cost share	Cost (in $)	Cost share	Cost (in $)	Cost share	(in $)
1	13410	90	247	2	389	3	872	6	14919
2	11097	93	124	1	355	3	385	3	11961
3	9005	67	2283	17	482	4	1579	12	13348
4	5363	92	56	1	316	5	121	2	5855
5	6199	80	493	6	80	1	960	12	7732
6	6017	91	98	1	382	6	81	2	6578
7	1806	91	NA	NA	176	9	NA	NA	1982
8	8788	69	647	5	1651	13	1636	13	12722
9	1891	32	1869	32	27	0	2119	36	5905
10	6699	98	NA	NA	116	2	NA	NA	6816
11	9851	88	97	1	779	7	443	4	11169
12	2755	96	30	1	76	3	4	0	2865
13	16673	99	NA	NA	152	1	NA	NA	16825
15	4684	87	206	4	107	2	382	7	5379
16	8322	81	1100	11	74	1	781	8	10278
Total Cost	**112,558**	**84**	**7,250**	**5**	**5,163**	**4**	**9,362**	**7**	**134,333**
Average Cost	**7,504**		**604**		**344**		**780**		**8,956**

Table 2: Results from Logistic Regression Models Used to Generate Predicted Outcomes

		Improvements in Behavior		Improvements in Self-efficacy	
Variables	**Variable Definition**	**Coef.**	**SE**	**Coef.**	**SE**
Gender	0 = Female; 1 = Male	0.061	0.238	-0.425*	0.233
Age	Age in years	0.036	0.110	-0.121	0.102
Ethnicity	1 = African American; 2 = White; 3 = Others	0.077	0.163	0.120	0.153
Family Type	0 = Single Parent; 1 = Both Parent; 2 = Others	0.101	0.197	0.068	0.188
Location	1 = Rural; 2 = Subruban; 3 = Cities	-0.754***	0.211	-0.468***	0.171
Class size	Number of students in a class	-0.022***	0.008	0.000	0.005
Group Type	0 = Single Grade; 1 = Mixed Grade	-0.995**	0.470	-0.632*	0.363
Volunteers	Number of volunteers assisting the class	0.987***	0.341	0.134	0.217
Duration	Number of days between pretest and posttest	-0.007***	0.002	0.007**	0.003
Handouts	Number of handouts given to the participants	0.117***	0.044	-0.009	0.031
PA Age	Age of PA in years	-0.071***	0.020	-0.002	0.013
PA Experience	Experience of PA in years	-0.042	0.039	-0.017	0.047
Constant		7.137***	1.587	4.882***	1.352
LL		**-294.187**		**-313.256**	
Pseudo R2		**0.092**		**0.069**	

** Significant at 10%; ** Significant at 5%;*** Significant at 1% levels.*

Table 3: Effects of Youth EFNEP by County

County ID	Number of participants		% Evaluated	Change in Behavior		Change in Self-efficacy	
	Reached	Evaluated		Observed	Predicted	Observed	Predicted
1	378	296	78	295	294	285	294
2	325	274	84	263	265	264	273
3	314	260	83	247	251	259	259
4	102	86	84	80	84	81	85
5	33	21	64	20	21	20	21
6	321	230	72	207	226	209	228
7	97	82	85	82	82	80	82
8	120	103	86	103	102	101	102
9	26	11	42	11	11	10	11
10	60	38	63	38	38	36	38
11	434	246	57	242	241	239	245
12	52	30	58	30	29	29	30
13	148	93	63	83	91	79	92
15	104	55	53	46	50	52	55
16	52	39	75	39	35	38	39
Total Effect	**2,566**	**1,864**	**73**	**1,786**	**1,819**	**1,782**	**1,853**
Average Effect	**171**	**124**		**119**	**121**	**119**	**124**

Table 4: The Estimates of Cost Effectiveness Ratio's by County

County ID	Cost Effectiveness Ratio (in $) per				
	Student Reached	Change in Behavior		Change in Self-efficacy	
		Actual	Predicted	Actual	Predicted
1	39	51	51	52	51
2	37	45	45	45	44
3	43	54	53	52	52
4	57	73	70	72	69
5	234	387	370	387	372
6	20	32	29	31	29
7	20	24	24	25	24
8	106	124	125	126	124
9	227	537	537	591	540
10	114	179	181	189	180
11	26	46	46	47	46
12	55	95	97	99	96
13	114	203	186	213	183
15	52	117	108	103	99
16	198	264	293	270	264
Total CER	**52**	**75**	**74**	**75**	**73**
Average CER	**89**	**149**	**148**	**153**	**145**

Table 5: Derivation of Confidence Interval for the CERs using Bootstrap

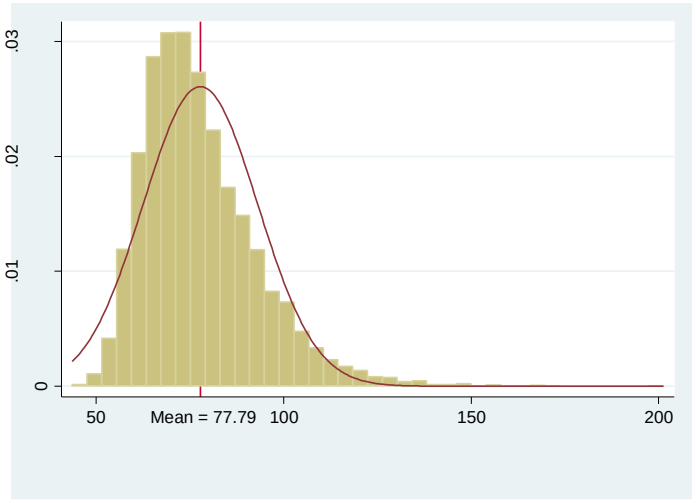
(Replications = 10000)

Outcomes	CER		Bias	SE	[95% CI]*
	Observed	Bootstrap			
Behavior (Actual)	75.21	77.79	2.58	15.29	$56.04 to $115.26
Behavior (Predicted)	73.83	76.37	2.53	15.07	$54.85 to $113.57
Self-efficacy (Actual)	75.38	78.01	2.62	15.55	$56.18 to $116.11
Self-efficacy (Predicted)	72.50	74.95	2.45	14.69	$53.99 to $111.35

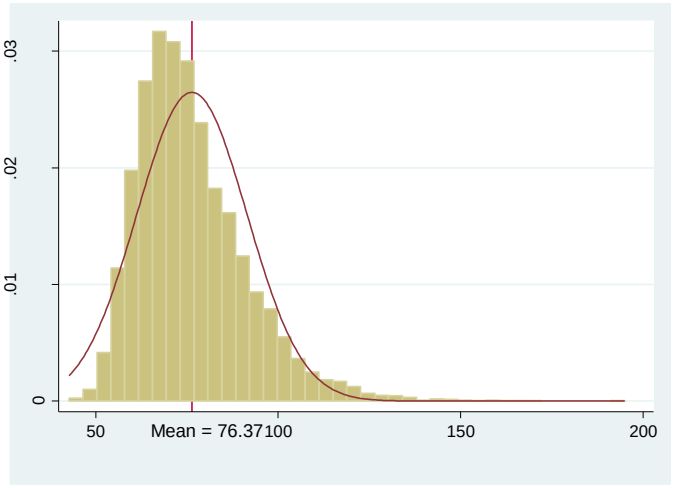
* *CI is estimated using the Bias-corrected method.*

Table 6: CER by Improvement Threshold

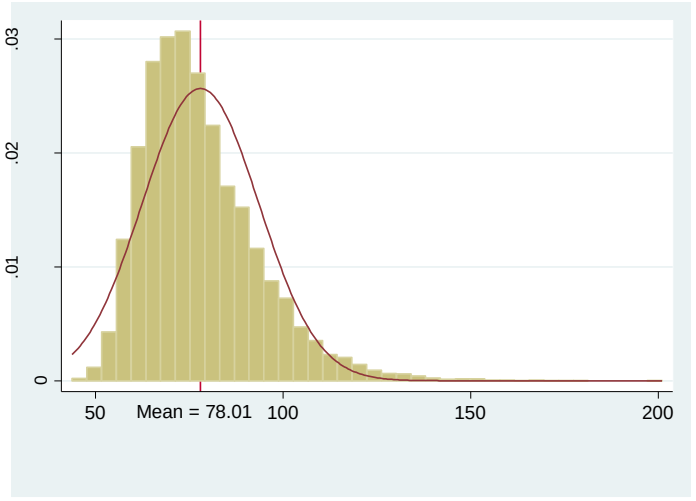
Thresholds on Improvements	CER per Behavior Improvement	CER per Self-efficacy Improvement
1	75	75
2	78	82
3	84	96
4	94	119
5	114	156
6	148	226
7	209	340
8	306	551
9	452	933
10	726	1658



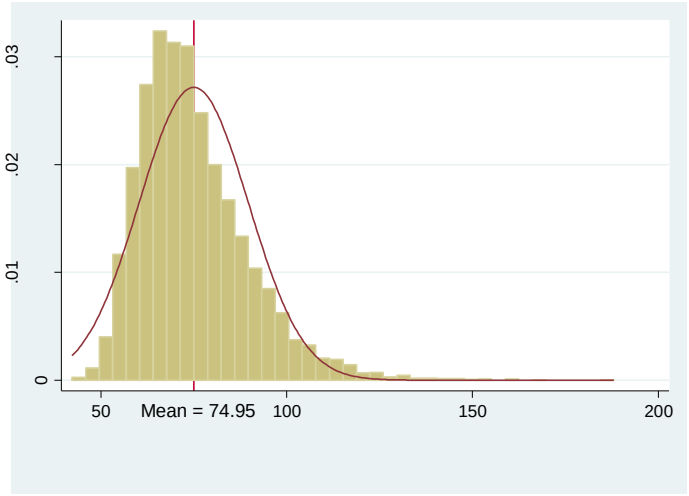
(a)



(b)



(c)



(d)

Figure 1. Histograms of Bootstrap Replicates of the Cost Effectiveness Ratio Estimates, Overlaid by the Normal Density of Same Mean and Variance.

Figures (a) and (b) use actual and predicted improvements in behavior as effects measure, respectively. Figures (c) and (d) use actual and predicted improvements in self-efficacy as effects measure, respectively.

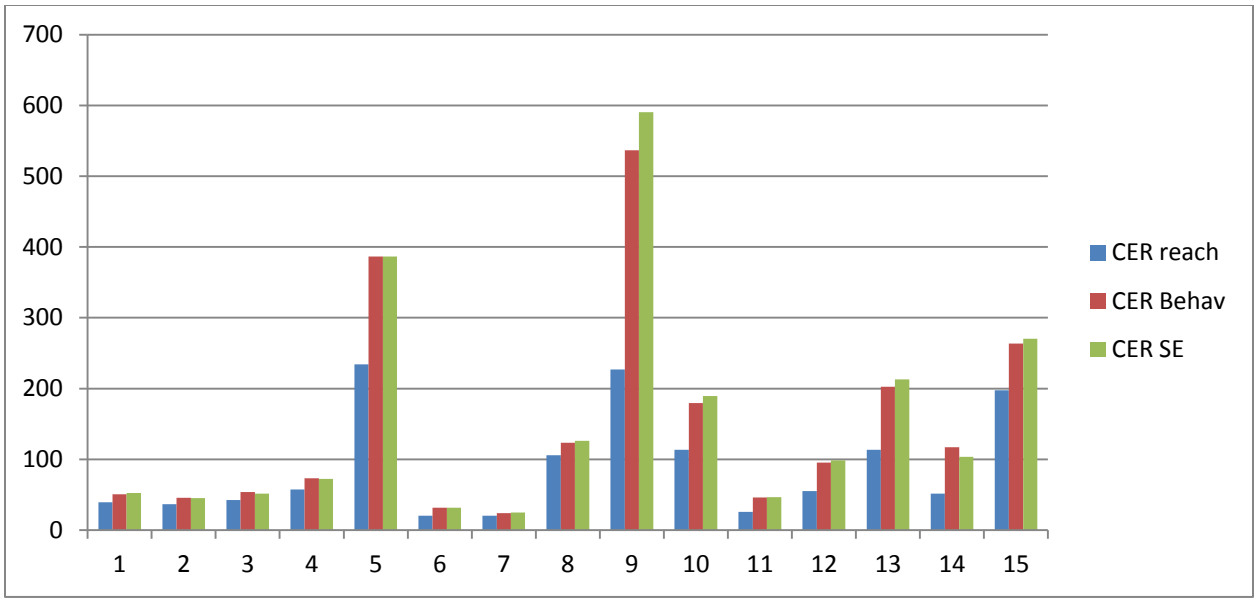


Figure 2. Cost Effectiveness Ratios by County

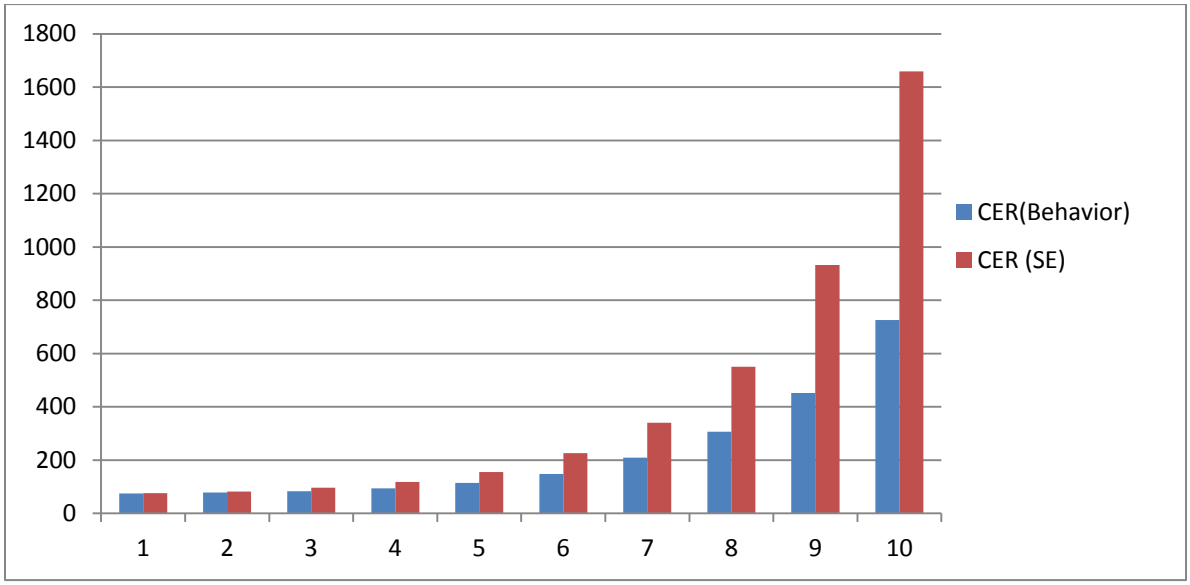


Figure 3. Cost Effectiveness Ratios by Threshold on Improvements

Appendices

Appendix.1: List of Members in Advisory Committee

1. Elena Serrano
2. George Davis
3. Gary Skaggs
4. Kathy Hosig
5. Marilyn Townsend
6. Mary McFerren
7. Michael Ellerbock
8. Michael Lambur
9. Nancy Franz
10. Peter Muennig
11. Susan Baker

Appendix.2: Development of Instrument for Measuring Effects

A.2.1. Preliminary Instrument

Table 7: Pool of Items in Preliminary Instrument

Items begin with the following:

Self-efficacy instrument: Each day, I think I can eat*/drink*…
Behavioral intention instrument: Each day, I plan to eat/drink …
Behavior instrument: Yesterday, I ate/drank …

** In the actual questionnaire, the word "eat" was used for solid food items and the word "drink" was used for liquid. For items related to physical activity, these terms were deleted.*

Domain	Item (I) #	Item description
Breads, Cereals, Grains	I1	… breads, grains and cereals
	I2	… whole grains Whole grains include whole wheat (breads, oatmeal, cornmeal, cornbread, brown rice)
	I3	… cold, sweetened cereal (Cold, sweetened cereal includes Frosted Flakes™, Froot Loops™, Apple Jacks™ etc)
Meat and Beans	I4	… meat
	I5	… kidney beans, or black beans, or brown beans, or red beans (Do not include green beans.)
	I6	… hamburger meat, hot dogs, sausage (chorizo), steak, bacon, bologna, or ribs
	I7	… fried meat with a crust

		(Fried meat with a crust includes chicken nuggets, fried chicken, chicken fried steak, fried pork chops, or fried fish.)
Milk	I8	… milk Include milk substitutes like soy milk, almond milk, and rice milk.
	I9	… fat-free or low fat milk (Include milk substitutes. Consider any flavor, smoothies, coffee drinks, and other beverages. Low fat milk is 1% and fat-free is skim milk or 0% milk, and they usually have green tops. *)*
	I10	… whole milk (Include milk substitutes. Consider any flavor, smoothies, coffee drinks, and other beverages. Reduced fat is 2%, and usually has blue tops. Whole milk usually has red tops.)
Fruits	I11	… fruit (Fruit includes pieces of fresh fruit, dried fruit, or 100% fruit juice.)
	I12	… *fresh* fruit
	I13	… *canned* fruit in syrup
Vegetables	I14	… vegetables
	I15	… vegetables that were orange, red, yellow, or green
Calorie dense food	I16	… fried (white) potatoes (Fried (white) potatoes include French fries, potato chips, and tater tots.)
	I17	… baked foods instead of fried foods (Baked foods include chicken without skin and pretzels.)
Desserts	I18	… dessert (Dessert includes cookies, cake, pie, or candy…)
Sodium	I19	… salty foods (Salty foods include cheese, crackers, potato chips, or bacon.)
	I20	… add salt to my food
Restaurants/Fast food	I21	… at fast food restaurants
	I22	… large or super-size portions at restaurants
Satiety	I23	… stop eating when I am full
Beverages	I24	… regular soda (Regular soda includes Coca Cola™, Pepsi™, and Mountain Dew.™)
	I25	… sports drinks (Sports drinks include Gatorade™ and Powerade.™)
	I26	… water
Knowledge on MyPyramid	I27	I know the food groups in MyPyramid
	I28	… according to MyPyramid
Physical Activity/ Sedentary Behavior	I29	… be physically active for _____ hours (Physically active includes running, jumping, and dancing.)

	I30	… watch TV, play video games, or use the computer for _____ hours
	I31	… how many days do you ….breathe hard for 30 minutes or more while physically active?

Table 8: Item Response Type

Response coding	Response format			Total score
	Four points Likert-type scale	**Two points scale**	**Five points Likert-type scale**	
Reverse coding	I3, I4, I6, I7, I10, I13, I16, I18, I19, I20, I21, I22, I24, I25, I30 [15 items]	[0 item]	[0 item]	15 items
No reverse coding	I1, I2, I5, I8, I9, I11, I12, I14, I15, I17, I23, I26, I28 [13 items]	I27 [1 item]	I29, I31 [2 item]	16 items
Total	28 items	1 item	2 items	31 items

Table 9: Scoring Rule for Items

Response format	Response	Code	Reverse code
Five points Likert-type scale	4 days (2 or more hours)	4	NA
	3 days (1 and half hour)	3	NA
	2 days (1 hour)	2	NA
	1 day (Half hour)	1	NA
	0 day (0 hour)	0	NA
Four points Likert-type scale	3 or more times a day (Strongly agree)	3	0
	2 times a day (Agree)	2	1
	1 time a day	1	2

	(Disagree)		
	No, I cannot (Strongly disagree)	0	3
Two points scale	Yes, I can	1	NA
	No, I cannot	0	NA

A.2.2. Item Analysis on preliminary instrument

A.2.2.1: Illustration for computing item total correlation

Tables A and B illustrate the computation of item-total correlation. A sample data matrix consisting of an instrument with five items and test scores from 10 individuals are used.

Table 10: Sample Data Matrix

Individual	Item 1 (A)	Item 2 (B)	Item 3 (C)	Item 4 (D)	Item 5 (E)	Individual total score (F)
1	3	1	0	0	0	4
2	1	0	2	1	0	4
3	3	3	0	3	3	12
4	1	2	2	3	0	8
5	3	3	0	2	1	9
6	0	1	3	2	0	6
7	2	1	3	2	0	8
8	3	2	0	0	2	7
9	2	2	2	2	1	9
10	3	3	3	3	0	12
Item total score	21	18	15	18	7	

Table 11: Computation of Item-Total Correlation

Individual	Individual total score -Item 1 (F-A)	Individual total score -Item 2 (F-B)	Individual total score -Item 3 (F-C)	Individual total score -Item 4 (F-D)	Individual total score - Item 5 (F-E)
1	1	3	4	4	4
2	3	4	2	3	4
3	9	9	12	9	9
4	7	6	6	5	8
5	6	6	9	7	8
6	6	5	3	4	6
7	6	7	5	6	8
8	4	5	7	7	5
9	7	7	7	7	8
10	9	9	9	9	12
Item- total correlation	0.047891	0.72115	-0.40196	0.480171	0.005405

The item-total correlation for each item is given in Table B. A low item-total correlation implies that the individuals who scored high in the test scored less in a given item.

A.2.2.2. Results from Item Analysis

Table 12: Results from Item Analysis on Self-Efficacy Instrument
Iteration 5: N items = 11, N individuals = 129, Cronbach's alpha = 0.8197

Item	Item description	Mean	SD	Item total correlation
SE3	Each day, I think I can eat cold, sweetened cereal	1.5659	1.1309	0.3881
SE6	Each day, I think I can eat hamburger meat, hot dogs, sausage (chorizo), steak, bacon, bologna, or ribs	1.1163	1.0049	0.5417
SE7	Each day, I think I can eat fried meat with a crust	1.6822	1.1591	0.4421
SE16	Each day, I think I can eat fried (white) potatoes	1.6667	1.0631	0.4967
SE18	Each day, I think I can eat dessert	1.4651	1.1114	0.5816
SE19	Each day, I think I can eat salty foods	1.7597	0.9745	0.565
SE20	Each day, I think I can add salt to my food	2.2248	0.8948	0.3841
SE21	Each day, I think I can eat at fast food restaurants	1.7907	1.1227	0.5074
SE22	Each day, I think I can eat large or super-size portions at restaurants	2.1938	1.0831	0.4828
SE24	Each day, I think I can drink regular soda	1.5581	1.172	0.6019
SE30	Each day, I think I can watch TV, play video games, or use the computer for _____ hours.	1.3101	1.0518	0.3637

Table 13: Results from Item Analysis on Behavior Intention Instrument
Iteration 6: N items = 11, N individuals = 110, Cronbach's alpha = 0.7993

Item	Item description	Mean	SD	Item total correlation
BI3	Each day, I plan to eat cold, sweetened cereal	1.8	1.1071	0.3218
BI6	Each day, I plan to eat hamburger meat, hot dogs, sausage (chorizo), steak, bacon, bologna, or ribs	1.2727	1.1244	0.5141
BI7	Each day, I plan to eat fried meat with a crust	1.7455	1.0873	0.5124
BI16	Each day, I plan to eat fried (white) potatoes	1.4455	1.1931	0.4021
BI18	Each day, I plan to eat dessert	1.2	1.1398	0.4025
BI19	Each day, I plan to eat salty foods	1.5	1.1067	0.5233
BI20	Each day, I plan to add salt to my food	2.1909	1.0536	0.4652
BI21	Each day, I plan to eat at fast food restaurants	1.7091	1.0077	0.6178
BI22	Each day, I plan to eat large or super-size portions at restaurants	2.3	0.9726	0.4554
BI24	Each day, I plan to drink regular soda	1.2364	1.1803	0.5336
BI30	Each day, I plan to watch TV, play video games, or use the computer for _____ hours.	1.0182	1.0403	0.303

Table 14: Results from Item Analysis on Behavior Instrument
Iteration 7: N items = 11, N individuals =261, Cronbach's alpha = 0.8008

Item	Item description	Mean	SD	Item total correlation
B1	Yesterday, I ate breads, grains and cereals	1.5632	1.0344	0.3529
B2	Yesterday, I ate whole grains	1	1.0077	0.3687
B8	Yesterday, I drank milk	1.4828	1.1422	0.4722
B9	Yesterday, I drank fat-free or low fat milk	0.9693	1.0373	0.4598
B11	Yesterday, I ate fruit	1.2874	1.1426	0.5574
B12	Yesterday, I ate *fresh* fruit	0.9885	1.0579	0.5209
B14	Yesterday, I ate vegetables	0.8851	0.9659	0.5804
B15	Yesterday, I ate vegetables that were orange, red, yellow, or green	0.7701	0.9161	0.5855
B17	Yesterday, I ate baked foods instead of fried foods (Baked foods include chicken without skin and pretzels.)	0.6015	0.8827	0.4382
B26	Yesterday, I drank water	1.9425	1.1571	0.4473
B31	Last week, how many days did you breathe hard for 30 minutes or more while physically active?	2.0651	1.581	0.3883

A.2.3. Results from Exploratory Factor Analysis on Preliminary Instrument

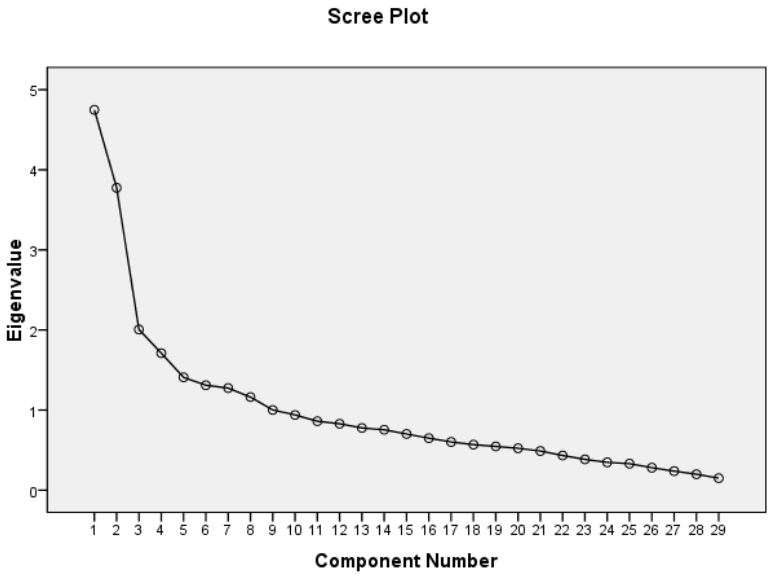


Figure 4. Scree Plot for Self-Efficacy Instrument

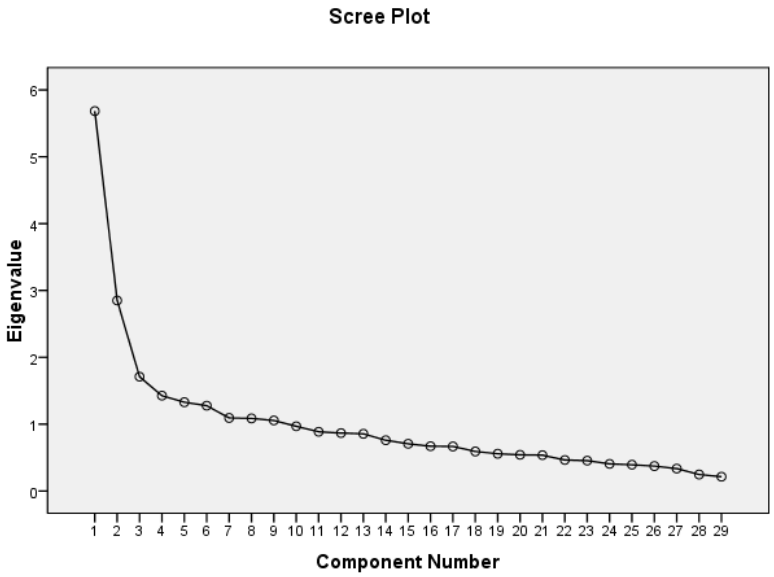


Figure 5. Scree Plot for Behavior Instrument

Table 15: EFA Factor Loadings using Oblique Rotation

Self-efficacy Pattern Matrix			Behavioral Intention Pattern Matrix			Behavior Pattern Matrix		
	Factor			Factor			Factor	
Questions	1	2	Questions	1	2	Questions	1	2
se6rc	.711	-.078	bi14	-.773	.187	b15	.735	.131
se24rc	.705	.343	bi15	-.671	-.128	b11	.684	.068
se18rc	.668	.180	bi10rc	.648	-.005	b14	.676	-.046
se19rc	.654	.060	bi11	-.600	-.155	b12	.607	-.038
se16rc	.611	.003	bi8	-.586	.052	b8	.568	-.054
se7rc	.588	-.147	bi12	-.583	-.039	b26	.564	.042
se21rc	.587	.133	bi5	-.567	-.082	b9	.541	-.030
se22rc	.581	.066	bi1	-.517	-.017	b1	.512	.156
se25rc	.502	-.240	bi9	-.513	.073	b10rc	-.474	.068
se20rc	.481	.144	bi13rc	.464	.265	b17	.442	-.251
se3rc	.466	-.075	bi31	-.391	-.040	b2	.416	-.063
se4rc	.445	-.178	bi2	-.369	-.018	b31	.414	-.090
se30rc	.437	.236	bi29	-.336	-.141	b6rc	-.381	.318
se17	-.391	.381	bi26	-.266	.052	b29	.380	-.006
se13rc	.258	-.173	bi21rc	-.047	.752	b4rc	-.365	.075
se14	.034	.691	bi24rc	-.038	.687	b13rc	-.363	.314
se15	-.126	.658	bi19rc	-.060	.679	b23	-.073	-.017
se1	-.230	.559	bi20rc	-.045	.632	b21rc	.015	.676
se11	-.066	.558	bi7rc	.246	.558	b22rc	-.093	.629
se8	.048	.552	bi18rc	-.124	.550	b24rc	.301	.608
se12	-.080	.492	bi6rc	.178	.540	b30rc	.367	.591
se10rc	.134	-.469	bi17	-.182	-.531	b19rc	-.058	.589
se9	-.033	.446	bi30rc	-.363	.525	b18rc	.086	.555
se5	.054	.422	bi16rc	.153	.513	b16rc	-.093	.546
se2	.100	.407	bi22rc	.288	.469	b20rc	-.086	.465
se31	.042	.362	bi25rc	.296	.360	b25rc	-.166	.465
se29	.068	.347	bi3rc	.205	.332	b7rc	-.149	.391
se26	.114	.338	bi4rc	.224	.296	b5	.265	-.333
se23	.042	.197	bi23	-.213	.238	b3rc	-.167	.245

Note on variables: SE refers to questions about self-efficacy, BI refers to Behavioral intention questions and B refers to behavior questions. RC represents the questions that were reverse coded because they were measuring negative dimensions. Numbers represents the question number.

Appendix.3: Final Instrument for Measuring Effects

HEALTHY WEIGHTS FOR HEALTHY KIDS EVALUATION

Before ☐ **After** ☐

Instruction: Please write in your answer to the following questions. When you respond, please print your answer.

Section 1: Tell us about yourself.

1. What is the **first letter** of your:

FIRST name _____ MIDDLE name _____ LAST name _____

2. Are you a: Boy ☐ Girl ☐ **[Please check one.]**

3. What **month** were you born? **[Please circle one.]**

January	February	March	April
May	June	July	August
September	October	November	December

4. What is your **favorite food**?

_____

5. What is your **favorite superhero**?

_____

6. How **old** are you? **[Please circle one.]**

7 8 9 10 11 12 13 14

Please circle your answer to the following questions:

	Questions	Response					
7	**How do you describe yourself?**	American Indian/ Alaskan Native	Asian	Pacific Islander	African American	White	Other
8	**Are you Hispanic or Latino?**	Yes			No		
9	**What language do you use with your parents most of the time?**	English	Spanish	Vietnamese		Chinese	Other
10	**Who do you live with?**	Mother	Father	Both mother and father		Grand parents	Other
11	**What grades do you usually get in school?**	Mostly A's	Mostly B's	Mostly C's	Mostly D's	Mostly F's or E's	

Section 2: What foods and drinks can you find at home?

Below are some questions about foods and drinks you had in your home last week.
This is *not* a question about eating, but just about *if* they were in your home. Please circle how many days last week the food or drink was in your home. **Circle just one.**

	Questions	How many days a week?								
12	Last week, we had **food made with whole grains** like whole wheat breads, cereals, brown rice, in our home _____ days a week	0 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	I don't know
13	Last week, we had **meat or kidney beans, or black beans, or brown beans, or red beans,** in our home_____ days a week	0 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	I don't know
14	Last week, we had **milk and milk products** like yogurt, cheese, in our home_____ days a week	0 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	I don't know
15	Last week, we had **fruit** like apple, oranges, cantaloupe, grapes, pears, in our home_____ days a week	0 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	I don't know
	Questions	How many days a week?								
16	Last week, we had **vegetables** like broccoli, green beans, spinach, corn, tomatoes, carrots, in our home_____ days a week	0 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	I don't know
17	Last week, we had **beverages** like Coke,™ Gatorade,™ fruit drinks, Snapple,™ in our home_____ days a week	0 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	I don't know

18	Last week, we had **desserts** like cookies, cake, candy or pies, in our home_____ days a week	0 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	I don't know
19	Last week, I had the chance to be **physically active at home** _____ days a week	0 days	1 day	2 days	3 days	4 days	5 days	6 days	7 days	I don't know

SELF-EFFICACY

Section 3: What do you THINK you can eat?

We want to ask you some questions about your *confidence* in eating different foods or drinking different beverages each day. We are *not* asking what you *actually* eat or drink or what you *should* eat or drink. This is not a test. There are no right or wrong answers. So, please be as honest as possible. Please read the 2 examples below:

	Example Questions	Times Per Day			
A.	**Each day, I think I can** eat a banana… I think I can eat a banana at breakfast, one for an afternoon snack, and one after dinner. So, that is a total of three times. I circle 3 or more.	0	1	2	3 or more
B.	**Each day, I think I can** eat carrots… I do not think I can eat any carrots. I circle 0.	0	1	2	3 or more

Please respond just like the examples above, by circling ONE number for each question.

Questions	Times Per Day

20	**Each day, I think I can** eat breads, grains and cereals….	0	1	2	3 or more
21	**Each day, I think I can** eat whole grains… Whole grains include whole wheat breads, oatmeal, cornmeal (cornbread), brown rice, and popcorn.	0	1	2	3 or more

Questions		Times Per Day			
22	**Each day, I think I can** eat kidney beans, or black beans, or brown beans, or red beans… Do not include green beans.	0	1	2	3 or more
23	**Each day, I think I can** drink milk… Include milk substitutes like soy milk, almond milk, and rice milk.	0	1	2	3 or more
24	**Each day, I think I can** drink fat-free or low fat milk… Include milk substitutes. Consider any flavor, smoothies, coffee drinks, and other beverages. Low fat milk is 1% and fat-free is skim milk or 0% milk, and they usually have green tops.	0	1	2	3 or more

25	**Each day, I think I can** drink whole milk… Include milk substitutes. Consider any flavor, smoothies, coffee drinks, and other beverages. Reduced fat is 2%, and usually has blue tops. Whole milk usually has red tops.	0	1	2	3 or more
26	**Each day, I think I can** eat fruit… Fruit includes pieces of fresh fruit, dried fruit, or 100% fruit juice.	0	1	2	3 or more

	Questions	Times Per Day			
27	**Each day, I think I can** eat *fresh* fruit…	0	1	2	3 or more
28	**Each day, I think I can** eat vegetables…	0	1	2	3 or more
29	**Each day, I think I can** eat vegetables that were orange, red, yellow, or green…	0	1	2	3 or more

30	**Each day, I think I can** eat baked foods instead of fried foods… Baked foods include chicken without skin and pretzels.	0	1	2	3 or more
31	**Each day, I think I can** drink regular soda… Regular soda includes Coca Cola,™ Pepsi,™ and Mountain Dew.™	0	1	2	3 or more
32	**Each day, I think I can** drink water…	0	1	2	3 or more

	Yes or No?				
33	I know what to eat for **MyPlate**...	Yes			No
	How much do you agree with this statement?				
34	**Each day, I think I can** eat meals and snacks according to **MyPlate**	Strongly Disagree	Disagree	Agree	Strongly Agree

Section 4: How physically active do you THINK you can be?

Below are statements about how active you think you can be. Please circle ONE category for each question. Be sure to read the question and response closely.

	Question	Hours Per Day				
35	**Each day, I think I can** be physically active for _____ hours. Physically active includes running, jumping, and dancing.	0	½	1	1 ½	2 or more

	Question	Days Per Week				
36	**Each week, how many days do you think you can** breathe hard for 30 minutes or more while physically active?	0	1	2	3	4 or more

BEHAVIOR

Section 5: What did YOU Eat or Drink Yesterday?

In this section, we want to ask you some questions about what you **actually** eat and drink – *NOT* what you *should* eat or drink or *think* you can eat or drink. There are no right or wrong answers. Please be as honest as possible.

	Example Questions	Times Per Day			
C.	**Yesterday, I ate a banana** Yesterday, I ate a banana at breakfast, at lunch, for an afternoon snack, and at dinner. SO, that is a total of four times. I circle 3 or more.	0	1	2	3 or more

D.	**Yesterday, I ate carrots…** Yesterday, I did not eat any carrots. I circle 0.	0	1	2	3 or more

Please respond just like the examples above, by circling ONE number for each question.

	Questions	Times Per Day			
37	Yesterday, I ate **breads, grains and cereals….**	0	1	2	3 or more
38	Yesterday, I ate **whole grains…** Whole grains include whole wheat breads, oatmeal, cornmeal (cornbread), brown rice, and popcorn.	0	1	2	3 or more

	Questions	Times Per Day			
39	Yesterday, I ate **meat…**	0	1	2	3 or more
40	Yesterday, I ate **hamburger meat, hot dogs, sausage (chorizo), steak, bacon, bologna, or ribs…**	0	1	2	3 or more
41	Yesterday, I drank **milk**… Include milk substitutes like soy milk, almond milk, and rice milk.	0	1	2	3 or more

42	Yesterday, I drank **fat-free or low fat milk…** Include milk substitutes. Consider any flavor, smoothies, coffee drinks, and other beverages. Low fat milk is 1% and fat-free is skim milk or 0% milk, and they usually have green tops.	0	1	2	3 or more
43	Yesterday, I drank **whole milk…** Include milk substitutes. Consider any flavor, smoothies, coffee drinks, and other beverages. Reduced fat is 2%, and usually has blue tops. Whole milk usually has red tops.	0	1	2	3 or more
44	Yesterday, I ate **fruit…** Fruit includes pieces of fresh fruit, dried fruit, or 100% fruit juice.	0	1	2	3 or more

	Questions	Times Per Day			
45	Yesterday, I ate ***fresh* fruit…**	0	1	2	3 or more
46	Yesterday, I ate ***canned* fruit in syrup…**	0	1	2	3 or more
47	Yesterday, I ate **vegetables…**	0	1	2	3 or more
48	Yesterday, I ate **vegetables that were orange, red, yellow, or green…**	0	1	2	3 or more

49	Yesterday, I ate **baked foods instead of fried foods…** Baked foods include chicken without skin and pretzels.	0	1	2	3 or more
50	Yesterday, I drank **regular soda…** Regular soda includes Coca Cola,™ Pepsi,™ and Mountain Dew.™	0	1	2	3 or more
51	Yesterday, I drank **water…**	0	1	2	3 or more
	How much do you agree with this statement?				
52	Yesterday**, I ate** meals and snacks according to **MyPlate**	Strongly Disagree	Disagree	Agree	Strongly Agree

Section 6: How physically active are YOU?

Below are statements about how active you are. Please circle ONE category for each question. Be sure to read the question and response closely.

	Question	Hours Per Day	
53.1	Yesterday, were you **physically active?**	Yes	No
If yes, *please answer question 53.2.* ***If No****, go to question 54.1.*			

53.2	Yesterday, I was **physically active for _____ hours.** Physically active includes running, jumping, and dancing.	½	1	1 ½	2 or more	
54.1	Yesterday, did you **watch TV, play video games, or use the computer?**		Yes	No		
If yes, *please answer question 54.2.* ***If No****, go to question 55.*						
54.2	Yesterday, I **watched TV, played video games, or used the computer** for _____ hours.		1	2	3 or more	
	Question		Days Per Week			
55	Last week, how many days did you **breathe hard for 30 minutes or more** while physically active?	0	1	2	3	4 or more

THANK YOU!

Appendix.4: Instrument for Measuring Costs

FORM_OS_12

Information on the Cost of Utilities, Value of Equipment, and Office Space for EFNEP Unit Offices <u>OWNED</u> by Local Government

Please consult with the appropriate local government fiscal officer to complete these items below. **Estimates are permitted, if made by someone who has information on which to base estimates.**

Please return this completed survey by September 20, 2012.

Section I: General Information

1. Date Form Completed: _____
2. Name and Title of Person Completing Form _____
3. Name of County Government Official(s) providing information: _____

   _____
4. Name of Extension Office out of which EFNEP Program Operates

   _____
5. Name of Building where Extension Office is Located _____
6. Street Address: _____
7. The building is OWNED by (Please check one):

   County ☐

   Town/City ☐

   State Government ☐

   Other_____

   Please Describe: _____

Section II: Information on the Cost of Utilities for Entire Office

Cost of utilities for the EFNEP unit office is needed for **September 2011 to May 2012**. If the cost is not available separately for the EFNEP unit within the building, please use the cost for the entire building where the Extension office is housed.

8. Cost of utilities reported below if for (Please check one):
   A. EFNEP unit office only ☐
   B. Entire building where EFNEP unit office is housed ☐
9. In the table below, please provide the actual cost of utilities

Cost Headings	**Amount in Dollars (Round to the Nearest $)**								
	2011				**2012**				
	Sep	**Oct**	**Nov**	**Dec**	**Jan**	**Feb**	**Mar**	**Apr**	**May**
Electricity									
Water									
Sewage									
Garbage									
Heating Oil/gas									
Janitorial or Maintenance Services									
Phone (including FAX line)[a]									
EFNEP Cell phone[a]									
Others (please list)									

a. For Phone (FAX) and Cell Phone these are the **monthly service charges**, not the actual cost of the hardware.

Section III: Information on the Cost of Equipment Used in Support of Youth EFNEP

In the table below, please provide information on the number and cost of equipment that are used by those significantly involved in **EFNEP.**
While we really just want equipment that is used by Youth EFNEP, we realize some items are shared by other programs in the office.

Equipment[a]	**Number**	**Amount in dollars per item ($)**	**Youth EFNEP Only**		**If Not Youth EFNEP Only, Estimate of Percentage of Time Used by Youth EFNEP**
			Yes	**No**	
Desktop Computer (Not Laptop)					%
Printers					%
Fax machines					%
Copiers (include price or yearly rent value and yearly maintenance fee)					%
VCRs					%
Camcorders					%
Digital Cameras					%
Others (Please Describe)					

a. This table refers to the actual **cost of the hardware**, not any monthly service charges.

Section IV: Information on Value of Office Space

10.	**Current Market Value** of building where EFNEP is housed.	**$**
11.	**Number of Extension Faculty/Staff** who have their PRIMARY office in the above space (Include only those with permanent positions).	#
12.	**Percentage** of above building occupied or used by Extension (Be sure to include meeting/conference rooms used frequently by Extension)	%
13.	**Percent of Extension's** space that is used (in anyway) for EFNEP staff (This should be based on the number of faculty/staff using VCE space and who are involved in EFNEP. Also include office space of other agents who give significant part of their time (more than 10%) to EFNEP. Also include space of conference and meeting rooms and storage of EFNEP materials, as well as common areas used by EFNEP staff)	%
14.	**Percent of EFNEP's** space that is used by Youth EFNEP Staff	%

Chapter 4: Evaluation of the Youth Expanded Food and Nutrition Education Program: What Contributes to its Effectiveness?

4.1. Introduction

The estimated cost effectiveness ratios (CERs) in the previous chapter provides a snapshot on how efficient the youth EFNEP program in Virginia is in achieving its stated program goals. There is however a wider variation in program outcomes across counties. An interesting follow up to the CERs, from a policy perspective, would be to understand the factors that contribute to the differences in effectiveness across programs. Identification of attributes of an effective program would help formulate the best practices for youth EFNEP thereby contributing to improve overall program effectiveness.

The purpose of this chapter is to explore determinants of the outcomes of the youth EFNEP program in Virginia. Specifically, this chapter aims to answers the following research questions:

1. Do student characteristics contribute or counteract the program's effectiveness in improving participant's nutrition behavior?
2. Do program characteristics contribute or counteract the program's effectiveness in improving participant's nutrition behavior?
3. Do program assistant characteristics contribute or counteract the program's effectiveness in improving participant's nutrition behavior?

Evaluation of the effectiveness of any program hinges upon how the effects or the outcomes of the program are measured (Rossi et al. 2004). Consistent with the goal of the youth

EFNEP program, the program outcomes in this study were defined as the improvements in participants' *nutrition behavior*. Outcome measurement of educational efforts such as the youth EFNEP is often complicated due to the various facets (dimensions) the educational programs ought to address. To comprehensively measure the outcomes no important dimension should be overlooked (Rossi et al. 2004). A comprehensive set of standardized instrument consisting of 19 questions were therefore developed to measure different dimensions of nutrition behavior. The previous chapter provides details about the process involved in developing this instrument and its content; thus it is not repeated here. The purpose of the instrument is to measure changes in nutrition behavior among program participants before and after the delivery of the curriculum.

The measure of change in outcomes, followed in the previous chapters, counted the number of individuals whose responses in any one of the items among the pool of 19 questions improved from pretest to posttest. Such measures which rely on improvements in any one question ignore the comprehensive depiction of nutrition behavior that the multiple item instruments aim to measure. Further, as we saw in the previous chapter that the choice of threshold on improvement can greatly influence the measure of effects. Despite some of these limitations, count based improvements measures were followed in previous chapters to be consistent with the current practice of the USDA in measuring EFNEP's outcomes (USDA 2013) and to make our results comparable to theirs. This chapter takes an alternative route to comprehensively measure the outcomes without overlooking any of the important dimensions of nutrition behavior.

Measurement issues related to multiple items are rife in social sciences, especially in education and psychology. One of the popular approaches that unify the responses to multiple items into a single composite measure is the Rasch measurement model. Rather than measuring

the individual items alone, the Rasch measurement model uses a more sophisticated probabilistic theory model to measurement (Fischer and Molenaar 1995). The underlying construct being measured is conceptualized as latent variable, and is thought of as being in a continuum. By modeling the individual's response to each item, the Rasch model generates individual specific scores which can be interpreted as their relative position on the latent trait (Bond and Fox 2007).

This chapter utilizes the Rasch measurement modeling framework to measure the outcomes of the youth EFNEP program among its participants. Application of the Rasch measurement model is novel in the current context, and is an improvement over the outcome measure in practice for the youth EFNEP program by the USDA. A similar measurement model has been used by the USDA to measure the food security for the US households (Bickel et al. 2000).

While obtaining a measure for the underlying latent variable is an improvement over just counting responses, ultimately there is interest in understanding what factors affect the latent construct. Traditionally, applications of the Rasch measurement models were limited to the measurement of individual specific effects, item specific effects, and some item differential functioning type analyses (Rijmen et al. 2003). More recently, the Rasch-type measurement models are conceptualized as broader classes of statistical models such as hierarchical linear models or multilevel models (Johnson and Raudenbush 2002; Raudenbush and Bryk 2002), generalized mixed effects models (Agresti 2002; Rabe-Hesketh et al. 2004), and multinomial models (Rijmen et al. 2003). The conceptualization of the Rasch-type measurement models within the general statistical modeling frameworks allows more flexibility for measurement and the inclusion of covariates. As we will see later, the basic Rasch-type measurement models can be extended to incorporate potential covariates and the repeated measures (pretest/posttest) data

(Johnson and Raudenbush 2002), which can be useful in explaining the inherent heterogeneity in the responses that are not explained by the simple Rasch-type models.

The chapter is organized as follows. Section 4.2 discusses the methods. The first part of this section describes a simple Rasch-type measurement model. What follows is the conceptualization of the basic Rasch-type model as broader classes of statistical/econometric models. Section 4.3 lays out the empirical model used in this chapter. Section 4.4 describes the data. Section 4.5 presents the results and discussion. Section 4.6 concludes with some discussion of the limitations. Introduction about the youth EFNEP program in Virginia and the data collection methods are not included in this chapter as they are already discussed in detail in the previous chapter.

4.2. Methods

Measuring Outcomes of Youth EFNEP: A Rasch Measurement Approach

The Rasch model is a simple but standard measurement model. It is based on the probability theory to measure latent traits (Rasch 1960). A series of items are defined in relationship to a single underlying latent variable that we want to measure. The probability of individual's response to these items is assumed to depend on the latent measure of the individual (called 'ability') and the difficulty of the item (called 'difficulty'). Individuals are believed to have different levels of latent abilities according to the number of questions they answer correctly in the test, and items are believed to have different levels of difficulties according to the number of individuals who answer those items correctly (Bond and Fox 2007; Smith and Smith 2004).

The Rasch model invokes several assumptions. One of the key assumptions of the Rasch model is *unidimensionality*, that is, the set of items included in the test should measure only one latent construct. This assumption ensures the additivity property which means that item difficulty parameters and individual ability parameters are additive in their effects, yielding a readily interpretable ordering of items and individuals on an interval scale (Rasch 1980). The additivity assumption also means that each item is equally discriminating, that is the slope of each item characteristic curve, which plots the probability of a correct response as a function of the person parameter, is identical. Other important assumption of the Rasch model is *local independence* which means that for a test of a given length, individual's item responses are conditionally independent given individual and item parameters. The *monotonicity* assumption ensures that the probability of a successful response is a monotone non-decreasing function of the latent trait parameter (Fischer and Molenaar 1995).

Consider administering a series of questions to measure a latent construct. According to the Rasch model, when an individual with some ability encounters an item of some difficulty, the probability of a correct response is modeled as a logistic function of the sum of ability and difficulty parameters (Fisher and Molenaar 1995). For items measured in a *dichotomous* response format, assuming independence of responses across individuals and items, the response probability for n^{th} individual on i^{th} item is modeled as:

$$Prob(Y_{ni} = 1) = \frac{\exp(\alpha_n + \beta_i)}{1 + \exp(\alpha_n + \beta_i)} \qquad (1)$$

where, n is a subscript for individuals ($n = 1,2, \dots, N$); i is a subscript for items ($i = 1,2, \dots, I$); Y_{ni} is the random variable representing the response of n^{th} individual to i^{th} item. Parameters α_n are simply individual effects and are interpreted as individual latent abilities. Parameters β_i are

item effects and are interpreted as item difficulty or item easiness.[13] The logistic model formulation in equation (1) allows the estimation of the parameters, α_n and β_i, independently of each other. For ease of discussion, it is useful to recognize and define the *latent index* as

$$\eta_{ni} = \alpha_n + \beta_i \qquad (2)$$

which is the log odds form of equation (1) and is simply a linear combination of person ability and item difficulty. When the responses are polychotomous ordered response with j categories ($j = 0,1, \dots, J$), the probability of responding to the j^{th} category of item i for individual n is generalized to

$$Prob(Y_{nij} = j) = \frac{\exp \sum_j (\alpha_n + \beta_{ij})}{\sum_k \sum_j \exp\,(\alpha_n + \beta_{ij})} \qquad (3)$$

This model for the ordered response is also called the Partial Credit Model (Masters 1982). This model gives rise to $(j-1)$ parameters for each item. In the case when the response categories are same across all items, a more parsimonious model can be formed, by imposing the restriction

$$\beta_{ij} = \beta_i + \tau_j \qquad (4)$$

where, β_i represents the overall difficulty level of item i and the additional parameter τ_j's are called thresholds or step difficulty parameters which represents the added difficulty involved in moving from category $(j-1)$ to j. The threshold parameters remain the same for all items. The resulting model is called the Rating Scale Model (Andrich 1978; Wright and Masters 1982) which is a special case of the partial credit model. Obviously substituting (4) into the latent index (2) just generalizes the item difficulty parameter.

[13] In the Rasch model, item parameters are typically modeled as item difficulty$(-\beta_i)$. When estimated as (β_i), the model will yield the reverse of item difficulty, called as item easiness.

In the Rasch model framework, the individual parameters α_n are associated with individual specific dummy variables and the item parameters β_i are associated with a set of item specific dummy variables. One of the issues with this simple specification of the Rasch model is that it ignores the potential factors that explain the differences in individual effects. All the variation in individual abilities is captured by individual dummy variables without actually modeling the possible heterogeneity. Though individual effects indicate the individual's relative position in the latent ability scale, it completely ignores the potential factors that lead to the differences in latent ability across individuals. When the research goal is to understand whether there are any factors driving the differences in individual effects, a simple Rasch model does not provide an answer. One way to deal with such quest is to follow a two-step procedure, where, in the first step, a simple Rasch model is estimated to obtain the individual specific effects. In the second step, a separate regression of the individual effects on potential covariates is conducted. Such two-step procedures are often criticized for not accounting for the exogenous influences in the first step which could induce a bias in the first step estimates that carries forward into the second step (Wang and Schmidt 2002).

Modeling individual effects as random parameter offers a solution to some of the problems inherent to the fixed parameter Rasch-type measurement models (Agresti 2002; Fischer and Molenaar 1995). Within this framework, the individual parameter α_n in equation 2 can be modeled as a function of a constant term and a stochastic component as follows

$$\alpha_n = \alpha_o + u_n \qquad (5)$$

The individual parameters α_n have a mean α_0 common to all individuals. The stochastic component u_n is assumed to be distributed as $u_n \sim N(0, \sigma_u)$, and uncorrelated with the item

indicators included in the model. The variation in the individual effect is captured by the variance of the stochastic component u_n. This formulation avoids the incidental parameter problem (Neyman and Scott 1948; Lancaster 2000) that arises by including n individual specific parameters for estimation. This formulation is also more flexible as it can be extended to add the observed variables $\boldsymbol{X}$ that might have generated the random individual parameter (α_n) into the model

$$\alpha_n = \alpha_o + \alpha_k \boldsymbol{X_k} + u_n \qquad (6)$$

This random parameter model formulation allows incorporating individual varying covariates in the model to explain the individual's position in latent ability continuum. It also provides an elegant way to address the two step estimation bias by modeling the individual heterogeneity in one step.

The extended formulation of the Rasch model or the random parameter model can be conceptualized as different types of statistical models, such as generalized mixed effects model (Agresti 2002; Rabe-Hesketh et al. 2004), and multilevel model (Johnson and Raudenbush 2002). One of the early applications of extending the Rasch model was done by Johnson and Raudenbush (2002). They viewed a generic Rasch model for a dichotomous response (equation 2) as a multilevel or hierarchical logistic regression model where items are nested within individuals[14] and specified equation 2 as a two-level logistic regression model. At the first level, the log-odds of the response probability are modeled as a linear function of the item indicators. In the second level, the level 1 intercept (α_n) is modeled as a function of a constant and individual specific random component. Johnson and Raudenbush (2002) further extended their

[14] Random parameter models can be viewed as multilevel or hierarchical models.

multilevel formulation to incorporate covariates that could potentially explain the individual effects, and also the repeated measures data for individuals over time. This is attained by allowing the individual specific random parameter at level 1 to be a function of explanatory variables at level 2 such as gender and by including the dummy variable to indicate two time periods. An example of their extended Rasch model is formulated below

Level 1: $\eta_{ni} = \alpha_n + \beta_{ni} Item_i$

Level 2: $\alpha_n = \alpha_0 + \alpha_1 Time + \alpha_2 Gender + u_n$

$\beta_{ni} = \beta_i,\ i = 1, \ldots I - 1$.

By combining level 1 and 2 equations we get

$$\eta_{ni} = \alpha_0 + \alpha_1 Time + \alpha_2 Gender + u_n + \beta_i Item_i \quad (7)$$

The individual ability estimates in equation 2 is essentially equivalent to $\alpha_n \approx (\alpha_0 + \alpha_1 Time + \alpha_2 Gender + u_n)$ in equation 7. The coefficient α_0 is the average ability of individuals at the base period. The coefficient on time dummy (α_1) measures the mean difference between individual ability estimates in two time periods. Similarly, the coefficient on gender α_2 captures the difference between individual parameters among boys and girls. The variance of random component u_n gives the variability in ability estimates across individuals. The random component is assumed to be uncorrelated with the all included variables. β_i is the item specific fixed parameter that captures item difficulty. In essence, the multilevel formulation can be further extended to more levels to account for the nested/hierarchical data structure.

4.3. Empirical Model

For empirical analysis, this chapter will extend the Johnson and Raudenbush (2002) slightly by utilizing the random parameter model specification of the form (in log odds form) as

$$\eta_{nij} = \alpha_n + \beta_{ij} Item_{ij}$$

$$\alpha_n = \alpha_o + \alpha_k \boldsymbol{X_k} + u_n$$

$$\beta_{ij} = \beta_i + \tau_j$$

$$\text{Or,} \quad \eta_{nij} = \alpha_o + \alpha_k \boldsymbol{X_k} + u_n + (\beta_i + \tau_j) Item_{ij} \qquad (8)$$

where, η_{nij} is the log odds of response for individual n on item i, response category j. The intercept of the model $\alpha_n$is a random parameter which is allowed to vary by individuals and is specified as a function of covariate vector X and a stochastic component u_n. The vector X includes the individual specific attributes such as age and gender; program specific characteristics and PA specific characteristics. The item parameters β_{ij} are specified as fixed parameters to be estimated that vary by item. For parsimony, the item parameters β_{ij} are restricted as $\beta_{ij} = \beta_i + \tau_j$. The threshold τ_j's are parameters to be estimated and are constant across all items.

Equation 8 is for a given point in time or for a single instrument set. The main objective of this paper is to study the change in individual effects and its attributes by using the repeated observation on each individual at two time periods (pretest and posttest). To achieve these objectives, data (items) from two time periods are pooled. To explain the individual effects α_n , the observed covariates related to individual characteristics are included in vector X. A dummy variable to indicate the pretest and posttest data is also included in X to capture the difference in

individual effect at the two time periods. In addition to the individual characteristics, the program and PA related characteristics such as class size and PAs experience are assumed to have an impact on the generation of the posttest individual scores. There is, however, no reason to believe that the program or PA related variables have any impact on the individual scores before the individuals are exposed to the program (pretest scores). To account for the effects of student, program and PA characteristics on posttest scores, all covariates related to the student, program and PA characteristics were interacted with the time indicator and were included as regressors in vector X. The fully specified empirical model is given below followed by the interpretation of model parameters.

$$\eta_{nitj} = \alpha_{nt} + (\beta_i + \tau_j) Item_{ij}$$

$$\alpha_{\text{nt}} = \alpha_{\text{o}} + u_n + \alpha_{Student} StudentCharacteristics + \delta_{Time}\, Time + \delta_{Time_{Student}} Time * StudentCharacteristics + \delta_{Time_{Program}} Time * ProgramCharacteristics + \delta_{Time_{PA}} Time * PACharacteristics$$

Or,

$$\eta_{nitj} = \alpha_{\text{o}} + u_n + \alpha_{Student} StudentCharacteristics + \delta_{Time}\, Time + \delta_{Time_{Student}} Time * StudentCharacteristics + \delta_{Time_{Program}} Time * ProgramCharacteristics + \delta_{Time_{PA}} Time * PACharacteristics + (\beta_i + \tau_j) Item_{ij} \quad (9)$$

The individual effect α_{nt} is allowed to be a function of various covariates and a stochastic term u_n. We assume that u_n are independent across individuals, and are distributed $u_n \sim N(0, \sigma_n)$. Also, $u_n$is assumed to be uncorrelated with the item indicators. The item

parameters β_i do not vary over individuals, and are constant over the two time periods. β_i can be interpreted as the overall difficulty of item i . The threshold parameters $\tau_j$are constant for all items. The coefficient on the time indicator gives the average difference between the individual effects from pretest to the posttest. α_0 is the average individual effect at the pretest. The coefficients on the interaction terms have a nice interpretation as the effects of a given variable on the change in individual effect. For example, if we consider a program characteristic variable, say class size, the coefficient on time-class size interaction term can be interpreted as the marginal effect of class size on the change in individual effect. The coefficients on the time interaction variables answer whether the included covariates have any impact on the change in individual effects.

Estimation of Model Parameters: One way to use category ordering in the polytomous response data is via modeling the cumulative odds ratio, equivalently called as a proportional odds model (Agresti 2002; Rabe-Hesketh et al. 2004), which is followed in this paper. The difficulty parameter on item 1 (β_1) and the average individual effect at pretest (α_0) both are set to zero for identification. Estimation of model parameter is done using marginal maximum likelihood estimation methods. Because the marginal likelihood does not have a closed form solution, parameters are estimated by numerically evaluating the marginal likelihood, using adaptive quadrature methods (Rabe-Hesketh et al. 2005).[15] Because the individual effects are integrated out of the marginal likelihood function and not estimated directly, the individual effects are obtained as predicted posterior means.

[15] The GLLAMM routine in Stata 11 was used for estimation (Rabe-Hesketh and Skrondal 2008).

4.4. Data

This paper uses survey data from all youth EFNEP participants in Virginia who participated in the HWHK program during school year 2011-2012. Data were collected from pre-post surveys administered to all participants in 15 counties of Virginia during September 2011 to May 2012. The survey consisted of 19 items related to nutrition behavior practices which are consistent with the Dietary Guideline for Americans 2005 and that cover the content of HWHK curriculum. Details on the instrument and its development are provided in chapter 3. Individuals' responses to all 19 items before (pre-test) and after (post-test) the curriculum are used to infer their nutrition behavior and the changes in nutrition behavior between two times. All items were measured on a 4 point Likert- type scale (0, 1, 2, and 3). [16] The interest is then what affected the change in individual's latent behavior scores.

The survey, in addition to nutrition behavior related items, also collected data on several demographic and household characteristics of participants. We hypothesize that participant characteristics such as age, gender, family type (single parent family) could have some effect on their existing nutrition behavior. Younger individuals may be less likely to change behavior after the program compared to the older individuals. Male participants could perceive the lessons differently than their female counterpart leading to different outcomes for male and female. Controlling for the demographics allow us to see the differences in individual effects by different socio-demographic characteristics.

Literature suggests that, besides knowledge about healthy and unhealthy foods, individual's eating behaviors depends on their taste and preference for food, and more

[16] Items 16, 17, and 19 were originally measured on a five point scale. For the purpose of analysis, these items were rescaled to a four point scale.

importantly on availability and accessibility of healthy foods (Domel et al., 1996, Hearn et al., 1998, Taylor et al., 2005, Bere and Klepp, 2005, O'Dea, 2003). These results are intuitive especially in the case of children who usually have less or no control over the availability and accessibility of food at home or school. They eat what they are offered and not necessarily what they know is healthy. Availability and accessibility of food is thus considered as a potential predictor of nutrition behavior. The survey also collected data on availability of food at participants' home in the last one week of survey. Students reported the number of days of a week for which a typical food (eight items) was available at their home. Types of food were divided into healthy (include food such as fresh fruits and vegetables) and unhealthy (include food such as soda) food. Food availability score one each for healthy and unhealthy food, were generated by adding the raw scores on the number of days healthy and unhealthy food were available. Unfortunately, the variable 'food availability' had many missing observations. About 40% of participants did not answer most of the availability related questions. The availability variable is therefore not used in the main regression analysis. A separate analysis of observations with complete data on availability was done, and the results are discussed briefly.

Although the program largely focuses on schools with limited resource children, there is a considerable heterogeneity across programs. All programs are managed and implemented by the youth EFNEP program assistants (PAs). The curriculum is provided in a school setting by educators, the youth PAs. Six lessons of HWHK are covered in six - one hour sessions that cover activities/experiences from each lesson. The exact contents or activities covered in each group/class, however, might vary across programs. Based on their assessment, individual PA's can choose among alternative activities/experiences within each lesson and customize the program depending upon class needs. Volunteers, usually class teachers, often accompany the

classes during program implementation. Depending on the convenience of school/class and PA's schedules, the duration of program and class size can vary substantially. Recognizing the variation in program characteristics across the programs, we hypothesize that the program attributes would have an impact on the improvements in individual behavior among its participants. In addition PA characteristics, which include their age and experience, also vary widely across PAs. We therefore hypothesize that PA specific characteristics has impacts on program outcomes.

Table 1 provides the summary statistics on the variables used in this paper. The effective sample consists of 1,842 participants whose pretest and posttest surveys were matched. The proportion of male and female participants in the sample is almost equal (mean = 0.49). The average age of the participant is about 9.7 years with a standard deviation of 1.1 years. Participants' age ranged from 7 to 14 years. About 30% of the participants belong to single parent household. The HWHK curriculum was delivered in 102 different groups/classes by PAs in various schools of 15 counties across Virginia. There were about 25 participants per class (range 3 to 104 participants). There were on average more than one volunteer (excluding the PA) who assisted during the curriculum delivery. Six lessons of the HWHK curriculum were completed, on average, in about 71 days. The range of program duration was very wide (3 to 231 days). About 38% of the programs were delivered in the rural farm areas and in rural areas with population less than 50,000. Regarding PA characteristics, the average age of PA was about 47 years (range 24 to 68 years). PAs, on average, had about 5.67 years of experience with the youth EFNEP program. Some PA had more than 18 years of experience whereas some others were less than 6 months into the program.

[Place Table 1 Approximately Here]

4.5. Results and Discussion

4.5.1. Measuring Outcomes of Youth EFNEP

The individual raw scores were generated by adding, for each individual, their responses on all items which provide a cursory measure of individual scores (effects). The maximum raw score possible for each individual is 57 and the minimum score possible is zero.[17] The average individual raw score was 29.57 in the pretest and it increased to 32.20 in the posttest.

A more elegant way to measure individual effects is to use a response probability model which measures individual effects as latent ability as specified in equation 9. Two models are estimated. In the first model, model 1, none of the attributes related to individuals, program and PA's are incorporated. The individual effects, however, are allowed to be different for pretest and posttest by including the time indicator. The second model, model 2, extends model 1 by incorporating individual specific covariates and the interaction terms (equation 9) to explain the individual effects. Model 1 is nested within model 2. Individual specific effects are not estimated directly as they are specified as a function of a stochastic term (u_n). For both models, the predicted values of individual effects at pretest (t = 0) and posttest (t = 1) were obtained as follows:

Model 1:

$\widehat{\alpha_{n0}} = \widehat{u_n}$, for t = 0

$\widehat{\alpha_{n1}} = \widehat{u_n} + \widehat{\delta_{Time}}\ Time$, for t = 1

Model 2:

[17] All items were measured as a 4-point scale (0, 1, 2, 3). The maximum score one can attain is 3 in one item, or 57 (3*19) in all 19 items.

$\widehat{\alpha_{n0}} = \widehat{u_n} + \widehat{\alpha_{student}}\, StudentCharacteristics$, for t = 0

$\widehat{\alpha_{n1}} = \widehat{u_n} + \widehat{\alpha_{student}}\, StudentCharacteristics + \widehat{\delta_{Time}}\ Time + \widehat{\delta_{Time_{Student}}}\, Time\ *$

$StudentCharacteristics + \widehat{\delta_{Time_{Program}}}\, Time\ * ProgramCharacteristics +$

$\widehat{\delta_{Time_{PA}}}\, Time * PACharacteristics$, for t = 1

The summary statistics of the predicted individual effects are given in Table 2. For both models 1 and 2, the average values of the predicted individual effects, both at the pretest and posttest are very close. For model 1, the average individual effect in the pretest is close to zero, and that at the posttest is 0.18 logits. Similarly, for model 2 the average pretest score is zero and the average posttest score is 0.19 logits. The average scores at the posttest give the average change in individual effects due to the program (0.18 logits for model 1 and 0.19 logits for model 2).

[Place Table 2 Approximately Here]

Figure 1 is a scatter plot of behavior scores for each individual at two periods. In the Y-axis are the pretest scores and in the X-axis are the posttest scores.[18] The line through the origin is a 45 degree line. All individuals who fall below the 45 degree line in the graph are the ones who scored higher at the posttest. There however are some individuals whose scores remain unchanged or declined at the posttest. These are the individuals who fall above the 45 degree line in Figure 2. The decline in behavior score for some individuals is somewhat counterintuitive but could be a consequence of the approach used to measure nutrition behavior. Nutrition behavior in our survey is measured by asking students what they ate the day before the survey, which might not represent the usual nutrition behavior practices for all participants. Measuring the

[18] The plots uses the transformed estimates of individual effects obtained from Model 2.

actual nutrition intake over a period of time (may be a week) would probably give a more precise measure of usual nutrition behavior. The tradeoffs with this approach, however, is the amount of resources needed to collect and analyze such data, not to mention the burden of keeping data/food records for young children.

The positive change in both the raw scores and the latent behavior scores among participants from pretest to posttest suggests positive impacts of the youth EFNEP program. However, the statistical significance of the change in scores can only be determined by looking at the model results which is done next.

[Place Figure 1 Approximately Here]

4.5.2. Factors Affecting the Outcomes of Youth EFNEP

Regression results from models 1 and 2 are presented in Table 3. Both models specify item parameters as fixed coefficients, and the individual effects as random parameters. In model 1, the random parameters are specified as a function of time indicator and a stochastic component u_n . Model 2 extends model 1 by incorporating the covariates related to individual characteristics. In addition, the program and PA characteristics related variables are interacted with time dummy to understand their impacts on change in outcomes from pretest to posttest. In both models, the random component u_n is assumed to be distributed normally with mean zero and variance σ_u. The estimated variance of u_n in model 1 is 0.788. The estimated variance of the random component is only slightly higher in Model 2 (0.800).

Result from model 1 show that the coefficient on the time dummy is positive and statistically significant meaning that the response probability is higher in posttest compared to pretest. The coefficients on all item parameters are statistically significant but not item 3,

implying that item 3 did not fit the data well. The coefficients on items can be interpreted as the overall item difficulty (rather easiness in the current case because they are estimated as β_i instead of $-\beta_i$). For difficult items, the individuals are less likely to choose higher values among the response options whereas for easy items individuals are more likely to choose higher values of response option. The coefficient estimate on item 10 (eating canned fruit) is smallest (-2.390), hence this item can be interpreted as the most difficult question. Item 16 (eating according to My Plate) on the other hand have the largest coefficient (1.671), and can be interpreted as the easiest item. Other items relatively easy for individuals to answer are 15 (of drinking water), and 17 (number of physically active hours). Items that are relatively difficult are 13 (eating baked food instead of fried), 12 (eating colored vegetables), 14 (reducing the intake of soda), 6 (consumption of low fat milk), and 7 (consumption of high fat milk). Estimates of relative item difficulty suggests that, for children, giving up unhealthy foods such as fried food, soda, and canned food is more difficult compared to other dimensions of nutrition considered here. Increasing the consumption of more healthy food such as colored vegetables is also relatively difficult. The preference for unhealthy food among children bolsters the need of nutrition education programs and puts an increased pressure on these programs to inculcate healthy behavior among its participant's. Nutrition education curriculum should focus more towards addressing the dimensions that children find difficult and to help children overcome the difficulties.

All items were measured on a 4-point scale. Thus there are three (J-1) threshold parameters estimated. Because the model is specified as a proportional odds model, the estimates of the threshold parameters (τ_j) are the cumulative log-odds of falling into or below category j. In other words, these parameters gives the log odds of responding to categories $<j$ versus all other categories$\geq j$. Specifically, τ_1 is the log odds for category 0 versus all other categories 1, 2,

and 3. Similarly, τ_2 is the log odds for categories 0, and 1 versus categories 2, and 3. Finally, τ_3 is the log odds for response categories 3 versus all other categories 0, 1, and 2. The thresholds are assumed to be constant for all items. The estimates of threshold parameters are almost equal for both models ($\tau_1 = -1.77;\ \tau_{2=} - 0.31;\ \tau_3 = 0.77$). These thresholds can also be interpreted as different intercepts for the sequence of log odds of (j-1) cumulative probabilities.

The results from model 2 are similar to the results from model 1. The coefficients on item dummy variables are all significant expect for item 3. Also, the magnitude of item coefficients remains largely unaltered. Inclusion of explanatory variables did not change the relative difficulty of items. The coefficient on time dummy in model 2 is positive. A joint significance test (F-test) of time dummy and all interaction terms (interacted with time dummy) in the model is jointly significant($\chi^2(10) = 414.89, P - value < 0.00$). The joint significance of all covariates involving time variable implies that the program has a significant effect in improving the response probability. This reinforces the findings from the previous model that the intervention has positive impacts on outcomes.

Of primary interest in this paper are the coefficients on the variables involving time interactions, which answers the effect of these variables on the change in nutrition behavior (change in individual effects). Among participants characteristics, participants' age and family type are significant predictors of change in outcomes. Participants' age has a negative impact on change in outcomes($\delta_{time\_age} = -0.031$). This result implies that the younger participants tend to perform better than the older ones. At a first glance this result is not quite intuitive because we would expect the older students to be more mature to understand the curriculum compared to the younger ones. However, if we look from a program perspective, the curriculum focuses on experiential learning principles and is developed to be age specific (Serrano et al. 2007). The age

specific design of the curriculum refutes the hypothesis that older participants might learn more. Our results that the program is more effective among younger participants bolster the early childhood interventions for cultivating healthy nutrition practices. Regarding participants' gender, whether a participant is male or female does not have a significant effect on behavior change, controlling for all other variables.

Another participant related variable, family type, has a positive and significant effect on change in behavior ($\delta_{time\_singlefamily} = 0.091$). Participants who belong to the both parent family (and/or who live with grandparents) secure better outcomes than the participants who belong to a single parent family, ceteris paribus. Other researchers have found that single parent households often have less income (Bianchi, 1995) and that such families tend to be more food insecure (Nord 2009). Despite the improved knowledge about healthy eating, participants from single parent family, may have less access to healthy food leading to smaller changes in their nutrition behavior. The corollary to this argument would be to assess the impact of food availability variable which is missing from the current analysis as a predictor. Literature has shown that availability and accessibility of food at home is an important predictor of nutrition behavior (Domel et al. 1996; Hearn et al. 1998; Taylor et al. 2005; Bere and Klepp 2005; O'Dea 2003). Because the program is designed to provide knowledge and to modify their preferences, food availability at home could have a positive effect on their nutrition behavior. Despite the forethought about this relationship and inclusion of this variable in data collection tools, a lot of participants did not provide complete data on food availability at home. This is understandable because the participants of this study are young children who might not have (accurate) information on food availability at their homes giving rise to missing data. About 40% of observations had missing data on food availability variable. Nevertheless, for exploratory

reasons, a separate analysis was conducted using only the sample with complete data on food availability variable. As expected, we found that the availability of healthy food at home was a highly significant predictor of outcomes. The significance of variable family type, however, became insignificant with the inclusion of food availability.

With respect to the program related covariates, the variables class size, number of volunteers, and program duration are significant predictors of change in outcomes. The variable class size has, perhaps not surprisingly, a negative impact on change ($\delta_{time\_classsize} = -0.006$), meaning that the participants who were taught in small classes have better outcomes than their counterparts taught in large classes. Class size varied widely across the programs. On average, there were 25 students in a class. Some classes had as few as 3 students, whereas some others had as many as 104 students. Small classes provide students more learning opportunity as the instructors could possibly reach to individual students. Such one-to-one interactions between instructors and participants might not be feasible in large classes, impeding student's learning.

The programs are often accompanied by volunteers (mostly class teachers). The numbers of volunteers has a positive impact ($\delta_{time\_volun} = 0.135$). The higher the numbers of volunteers who support in program delivery the better are the outcomes. More volunteers mean more learning opportunities for participants via personal interaction.

The variable program duration also had a positive impact on changes in outcomes ($\delta_{time\_duration} = 0.002$). This means that the participants who completed the curriculum in short duration had less change in outcomes than their counterparts who participated in long duration programs. Duration of program varied widely across the programs. PAs have a choice to space six – one hour lessons of the HWHK curriculum over a period of one

school calendar year. The average program duration was observed to be about 71 days (or 10 weeks). However, some programs were completed in as few as 3 days, whereas the others took 33 weeks to complete. Short duration programs performing worse off than the long duration programs could be due to the information overload in programs designed to complete in a very short duration.

Regarding the PA characteristics, PA's age has a positive and statistically significant effect on change in outcomes ($\delta_{time_{PAage}} = 0.003$). Experience of PA also has a positive and statistically significant effect on change in outcomes ($\delta_{time_{PAexper}} = 0.007$). The correlation between variables PA age and PA experience is not very high ($\rho < 0.5$). PA's with higher years of experience could have a better understanding of designing and delivering the curriculum more effectively. With higher experience, PA's might have mastered the tricks to better educate youths, leading to better outcomes. Positive relationship between PA's experience and the outcomes highlights the significance of collaborative experience sharing meetings among PAs. Sharing experiences might help PAs to learn from each other's experience and help improve the overall effectiveness of the program.

[Place Table 3 Approximately Here]

4.6. Conclusion and Limitations

This paper investigates the outcomes and effectiveness of the youth EFNEP program, in Virginia. A novel measure of outcomes used in this paper provides an alternative perspective to youth EFNEP evaluation. Current practice in evaluation of the youth EFNEP is based on counting the number of individuals whose response improves in one or more items between the

tests (USDA 2013). A Rasch-type measurement model is used in this paper which generates a latent behavior index for each individual by modeling individual's responses to all questions rather than only focusing on individual items. When the program aims to improve the overall behavior, an outcome measure that integrates the all dimensions of behavior seems to closely mimic the essence of policy objective.

The extended Rasch-type model which specifies individual effects as random parameters are more flexible than the simple Rasch model which specify the individual effects as fixed parameters. The former approach provides an opportunity to incorporate covariates to model the heterogeneity in individual effects. The usefulness of this research is also due the application of this approach to identify the factors affecting the change in individual behavior outcomes.

Overall, the youth EFNEP program in Virginia is effective in achieving its stated goals. The effectiveness of the youth EFNEP program is significantly predicted by individual specific characteristics, as well as program and PA characteristics. Young children tend to have better outcomes than the older ones. Effectiveness of programs among younger participants reinforces the significance of starting interventions to change nutrition behavior in early years of life. The importance of variable family type and food availability at home in explaining behavior change calls for increased coordination between USDA's food assistance programs and nutrition education programs like youth EFNEP.

The impact of program characteristics such as class size, number of volunteers, and program duration on program outcomes is an important finding. Programs implemented in small classes have better outcomes than the programs in large class. Programs which have more volunteers to assist in program delivery have better outcomes. Also, programs which were

spanned over relatively long duration had better outcomes than programs which completed in short duration. Significance of these program related variables would mean that the youth EFNEP program might benefit from providing some structure to the program delivery. Developing program operation guidelines to help PAs structure the program might be a strategy to consider for the youth EFNEP in order to maximize the program outcomes. Alternatively, these results also suggest that program effectiveness could be improved by allocating more resources into the program. Specifically, delivering program in small classes would mean that EFNEP would need more youth PAs and more volunteers to reach the same number of participants. Potential improvement in program effectiveness would therefore incur increase in program cost.

Both age and experience of PA have a positive and significant impact on program outcomes. Collaborative experience sharing meetings among PAs would provide an opportunity to discuss experiences and disseminate best practices between PAs that would help improve the overall effectiveness of the program. In addition, policies to retain and incentivize more experienced PAs would be an important aspect to consider for improving overall program effectiveness.

Some of the limitations are noteworthy. Effectiveness of the youth EFNEP program could potentially be influenced by other factors not considered in this paper. The youth programs differ in some aspects of program delivery, for example, some programs provide supporting materials such as teaching learning aids or handouts, to the participants. Different types of handouts are associated with each lesson and PAs choose to provide such materials to the participants as a supporting material to help improve learning effectiveness. The effectiveness of such teaching learning aids might translate into program effectiveness. A more important aspect

to consider would be the exposure of the participants to other nutrition related programs/lessons, either concurrent with the youth EFNEP lessons, or any previous exposures. Inability to control for such external influence might possibly over/under estimate the effects of the program. However, both the regression models specified in this study yield robust estimates of individual effects which increase confidence in the results we find, and suggest that the youth EFNEP program in Virginia is effective.

References

Agresti, A. 2002. Categorical Data Analysis. 2nd Edition. Wiley-Interscience publication. New Jersey.

Anderson, E. 1970. Asymptotic Properties of Conditional Maximum Likelihood Estimators. *Journal of the Royal Statistical Society,* 32: 283-301.

Andrich, D.1978. A Rating Formulation for Ordered Response Categories. *Psychometrika*, 43(4): 561–573.

Bere, E., and K. Klepp. 2005.Changes in accessibility and preferences predict children's future fruit and vegetable intake. *The International Journal of Behavioral Nutrition and Physical Activity*, 2:15.

Bianchi, S. M. 1995. The changing demographic and socioeconomic characteristics of single parent families. *Marriage and Family Review*, 20: 71–97.

Bickel, G., M. Nord., C. Price., W. Hamilton., and J. Cook. 2000. Guide to measuring Household Food Security, Revised (2000). U.S. Department of Agriculture, Food and Nutrition Service.

Bond, T.G., and C.M. Fox. 2007. Applying the Rasch Model: Fundamental Measurement in the Human Sciences, 2nd Edition. Mahwah, NJ: LEA publishers.

Chamberlain, G. 1980. Analysis of Covariance with Qualitative Data. *Review of Economic Studies*, 47: 225-238.

Domel, S.B., W.O. Thompson., H.C. Davis., T. Baranowski., S.B. Leonard., and J. Baranowski. 1996. Psychosocial predictors of fruit and vegetable consumption among elementary school children. *Health Education Research*, 11(3): 299-308.

Emberston S. E., and S.P. Reise. 2000. Item Response Theory for Psychologists. Mahwah, NJ: Lawrence Erlbaum Associates.

Fisher, G.H., and I.W. Molenaar. Eds. 1995. Rasch models: Foundations, Recent Developments and Applications. NY: Springer-Verlag.

Greene, W.H. 2005. Econometric Analysis. 5th Edition. Prentice Hall. New Jersey.

Hearn, M.D., T. Baranowski., J. Baranowski., C. Doyle., M. Smith., L.S. Lin., and K. Resnicow. 1998. Environmental Influences on Dietary Behavior Among Children: Availability and Accessibility of Fruits and Vegetables Enable Consumption. *Journal of Health Education*, 29(1): 26-32.

Johnson, C., and S. W. Raudenbush. 2002. A Repeated Measures, Multilevel Rasch Model with Application to Self-Reported Criminal Behavior. Research Report submitted to MacArthur Foundation and the National Institute of Justice.

Johnson, C., and S.W. Raudenbush. 2006. A Repeated Measures, Multilevel Rasch Model with Applications to Self-Reported Criminal Behavior. *Methodological Issues in Aging Research*. Cindy S, ed; Boker Steven M, ed. Norte Dame Series on Quantitative methods. Lawrence Erlbaum Associates Publishers, Mahwah, NJ.

Lancaster, T. 2000. The Incidental Parameter Problem since 1948. *Journal of Econometrics*, 95: 391-413.

Masters, G.N. 1982. A Rasch Model for Partial Credit Scoring. *Psychometrika*, 47: 149-174.

McFadden, D. 1974. The Measurement of Urban Travel Demand. *Journal of Public Economics*, 3:303-328.

Neyman, W., and E. Scott. 1948. Consistent Estimates Based on Partially Consistent Observations. *Econometrica*, 16: 1-32.

Nord, M. 2009. Food Insecurity in Households with Children: Prevalence, Severity, and Household Characteristics. EIB-56. U.S. Dept. of Agriculture, Econ. Res. Serv. September 2009

O'Dea, J.A. 2003. Why do kids eat healthful food? Perceived Benefits of and Barriers to Healthful Eating and Physical Activity among Children and Adolescents. *Journal of American Dietetic Association*, 103(4):497-501.

Rabe-Hesketh, S., A. Skrondal., and A. Pickles. 2005. Maximum Likelihood Estimation of Limited and Discrete Dependent Variable Models with Nested Random Effects. *Journal of Econometrics*, 128 (2): 301-323.

Rabe-Hesketh, S., and A. Skrondal. 2008. Multilevel and Longitudinal Modeling Using Stata. 2nd Ed. StataCorp LP, College Station, Texas.

Rasch, G. 1960. Probablisitc Models for Some Intelligence and Attainment Tests, University of Chicago Press (1980).

Raudenbush, S. W., and A. S. Bryk. 2002. Hierarchical Linear Models: Applications and Data Analysis Methods. 2nd ed. Advanced Quantitative Techniques in the Social Science Series 1, Sage publications, London.

Rijmen, F., F. Tuerlinckx., P. DeBoeck., and P. Kuppens. 2003. A Nonlinear Mixed Model Framework for Item Response Theory. *Psychological Methods*, 8(2): 185-205.

Rossi, P.H., M. W. Lipsey., and H. E. Freeman. 2004. Evaluation: a Systematic Approach. 7th Ed. Sage Publications. Thousands Oaks, California.

Serrano, E., S Diehl., K. Jamison., J. Shelhamer, J. Wages., and K. Gehrt. 2007. Healthy Weights for Healthy Kids: How Virginia Cooperative Extension is Addressing Childhood Overweight. *Food, Nutrition, and Health Publication,* Virginia Cooperative Extension, 348-275.

Smith, E.V., and R.M. Smith. Eds. 2004. Introduction to Rasch Measurement. Maple Groove MN: JAM Press.

Taylor, J.P., S. Evers., and M. McKenna. 2005. Determinants of Healthy Eating in Children and Youth. *Canadian Journal of Public Health*, 96.

Wang, H.J., and P. Schmidt. 2002. One-Step and Two-Step Estimation of the Effects of Exogenous Variables on Technical Efficiency Levels. *Journal of Productivity Analysis*, 18: 129-144.

Wright, B.D., and G. N. Masters. 1982. Rating Scale Analysis, Chicago: MESA Press.

United States Department of Agriculture (USDA).2013. National Institute of Food and Agriculture (NIFA).Expanded Food and Nutrition Education Program, EFNEP Impacts. Washington DC. Available at http://www.nifa.usda.gov/nea/food/efnep/impacts.html Accessed April 10, 2013.

Table 1: Summary Statistics on Explanatory Variables

Variables	Description	Mean	SD	Min	Max
Participant Characteristics *(N = 1842)*					
Gender	0 = Female; 1 = Male	0.49	0.50	0	1
Age*	Age in years	9.74 (9.94)	1.10 (1.10)	7	14
Family Type*	Whether participant live with 0 = Single parent; 1 = Others (Both parent; Grandparents/others)	0.71	0.44	0	1
Food* Availability Healthy	Number of days in the last one week in which healthy food was available (N=934 at pretest; N=1151 at posttest)	27.33 (28.32)	9.48 (9.44)	0 0	42 42
Food* Availability Unhealthy	Number of days in the last one week in which unhealthy food was available (N=1154 at pretest; N=1585 at posttest)	8.30 (7.90)	4.38 (4.48)	0	14
Program Characteristics *(N = 102)*					
Class Size	Number of students in a class in which the curriculum is taught	24.80	18.93	3	104
Volunteers	Number of Volunteers who provided assistance in delivering the program	1.31	0.62	0	4
Duration	Number of days between pretest and posttest	71.17	63.12	3	231
Residence	Location at which the program is delivered 0= Rural (farm, rural areas with population 10-50 thousand); 1= Suburban areas and central cities	0.38	0.48	0	1
Program Assistant Characteristics *(N = 15)*					
Age	Program Assistant's age in years	47.33	16.03	24	68
Experience	PA's experience with youth EFNEP (in years)	5.67	4.81	0.5	18

Table 2: Summary Statistics on Individual Effects

Model Type	Description	Mean	SD	Min	Max
Raw scores	Pretest Scores	29.57	9.34	0	54
	Posttest Scores	32.20	9.12	1	54
Model 1 predictions of Individual Effects	Pretest Scores	0.00	0.84	-2.61	3.41
	Posttest Scores	0.18	0.83	-2.42	3.60
Model 2 predictions of Individual Effects	Pretest Scores	0.00	0.84	-2.58	3.41
	Posttest Scores	0.19	0.85	-2.49	3.63

Table 3: Results from Regression Analysis

	Model 1		Model 2	
Variables	**Coef.**	**SE**	**Coef.**	**SE**
Itemdum2	-0.594***	0.042	-0.596***	0.042
Itemdum3	-0.021	0.042	-0.021	0.042
Itemdum4	-0.690***	0.043	-0.692***	0.043
Itemdum5	-0.273***	0.043	-0.273***	0.043
Itemdum6	-1.131***	0.043	-1.134***	0.043
Itemdum7	-1.253***	0.044	-1.255***	0.044
Itemdum8	0.203***	0.043	0.204***	0.043
Itemdum9	-0.186***	0.042	-0.187***	0.042
Itemdum10	-2.390***	0.047	-2.396***	0.047
Itemdum11	-0.652***	0.042	-0.653***	0.042
Itemdum12	-0.944***	0.043	-0.947***	0.043
Itemdum13	-1.067***	0.043	-1.070***	0.043
Itemdum14	-0.912***	0.044	-0.913***	0.044
Itemdum15	1.011***	0.045	1.014***	0.045
Itemdum16	1.671***	0.047	1.681***	0.047
Itemdum17	0.945***	0.044	0.948***	0.044
Itemdum18	-0.132***	0.042	-0.132***	0.042
Itemdum19	0.399***	0.044	0.401***	0.044
Time dummy	0.187***	0.015	0.094	0.172
Age			0.003	0.019
Gender			-0.011	0.044
Single family			-0.031	0.040
Time*Age			-0.031**	0.014
Time*Gender			0.024	0.029
Time*Single Family			0.091***	0.034
Time*Class Size			-0.006***	0.001
Time*Volunteers			0.135***	0.030
Time*Duration			0.002***	0.000
Time*Residence			0.045	0.041
Time*PA Age			0.003*	0.002
Time*PA Experience			0.007*	0.004
Random Effect				
σ_u	0.788	0.030	0.800	0.031
Threshold Parameters				
τ_1	-1.770***	0.038	-1.776***	0.191
τ_2	-0.310***	0.037	-0.311*	0.191
τ_3	0.776***	0.037	0.777***	0.191

Fit Statistics		
Log Likelihood	-84470	-84338
AIC	168986	168747
BIC	169196	169067
N	69236	69236

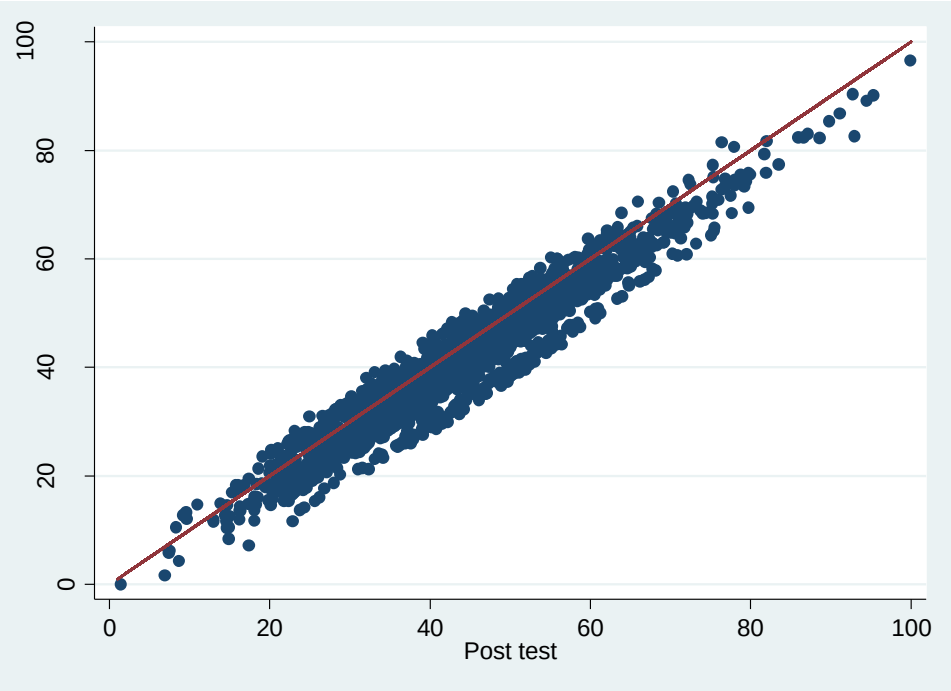


Figure 1. Scatter Plot of Individual Scores at Two Time Periods

Chapter 5: Conclusions

EFNEP is an ongoing effort of the USDA to inculcating the healthy nutritional practices among the low income population in the US. The US government commits millions of dollars every year for the same cause. The program has endured for more than half a century and has morphed, over the years, to serve the needs of the targeted audience throughout all states and US territories. The success of the EFNEP in achieving its stated goals can partly be witnessed in the annual impact reports published by the USDA and in several other independent researches published over the years. From an economic perspective, however, there is a lack of understanding with respect to some of the basic questions related to the EFNEP effectiveness and the allocations received by programs. This dissertation attempts to fill that gap by studying the costs and effectiveness of both adult and youth EFNEP programs. The scope and approaches engaged in the evaluation of the two programs are, however, different.

Chapter 2 of this dissertation examines the effectiveness of the adult EFNEP in achieving its goals at the national, regional, and state level. The returns to scale on three outcomes used by the USDA were estimated by using the Nutrition Education Evaluation and Reporting System (NEERS) data and the annual allocation data by state and territories for the years 2000–2006. The impacts of program and participant characteristics on the outcomes were also estimated. We find that after controlling for participant and program characteristics, the amount of money spent on the EFNEP has a positive and significant impact on two of the three outcome indices used by USDA. In general, there are constant and increasing returns to scale for two of the three federal outcome indices for most states, although not for all. In addition,

program and participant characteristics do not seem to be as important as the amount of money spent on the program in determining the program outcomes.

Chapters 3 and 4 are related to the evaluation of the youth EFNEP program. A general framework for conducting a cost effectiveness analysis for the youth program was developed in chapter 3, and as an application, the cost effectiveness analysis was done for the youth program in Virginia. The lack of instrument to measure the outcomes of the youth EFNEP was identified as one of the main reasons for the lack of evaluation studies on youth program. So, we developed and validated the effects instrument to measure the outcomes of youth EFNEP which are defined as the change in nutrition behavior, and change in nutrition self-efficacy among the program participants. Changes in outcomes were measured by employing the effects instrument at two time periods, pretest and posttest. On the costs side, only the direct costs associated with implementing the program were considered to reflect the value of resources used in the provision of the youth EFNEP-HWHK program in Virginia. The estimates of CER were based on data from 15 counties in Virginia which implemented the youth EFNEP-HWHK curriculum during the school year 2011 - 2012. The CER was estimated to be around $75 per behavior outcome, which means that the youth EFNEP program in Virginia spends on average about $75 to improve one nutrition related behavior per participant. The estimate of CER per improvement in self-efficacy is also about $75 per participant. Cross county comparisons reveal a wide variation in the estimates of CER across counties. These estimates of CER are the first of its kind for the youth EFNEP program, and no inference about the relative cost effectiveness can be yet made, however, we can conclude that the youth program in Virginia is more cost effective compared to the adult program (Baral et al. 2013). In the future, the tools and procedures developed in this study can be utilized by other youth EFNEP programs to generate information on the cost

effectiveness for respective programs. Results then can be used by the state and national EFNEP administrators to demonstrate program success, for comparing the cost effectiveness across states.

Chapter 4, as a follow up to chapter 3, specifically analyzes the outcomes of the youth EFNEP and investigates the factors contributing to the effectiveness of the program by utilizing the data from Virginia youth EFNEP. A noble approach to measurement of outcomes was followed in chapter 4. Unlike the usual practice of the USDA in measuring outcomes of EFNEP, which is to count the number of individuals who improve in at least one question, we consider nutrition behavior as a latent construct and used a Rasch-type measurement model to measure the outcomes of the youth EFNEP program. We find that on average the latent behavior scores for participants were higher at the posttest compared to that at the pretest, indicating the success of program in achieving the stated goals of nutrition behavior change. We also found that several characteristics related to the participant (age, family type), program (class size, number of volunteers, program duration) and PA (age and experience) are important attributes of effective program. Understanding the attributes of effective program is useful for EFNEP administrators as it can provide guidance for devising more effective programs.

This dissertation is not without limitations. Some of the limitations are already mentioned in the respective chapters. A few of which apply broadly to the current practice of the USDA for measuring the EFNEP effectiveness as well as measures used in part of this dissertation is worthy of further discussion.

At the pivot of any program evaluation are the measures of program outcomes. The outcomes used in chapter 2 and 3 of this dissertation follows the USDA measure of

improvements which, for each individual, are counted as whether they improved on *at least one item* among a list of items. We realize that the *improvement threshold,* which can be defined as *the number of items an individual improves on,* is an important variable that would affect the results. The improvement threshold could be any number between 0 and the total number of questions in the survey. That is, suppose there are three items (questions) on the survey. One could calculate the number of people who improved on 0 questions, 1 question, 2 questions, or all 3 questions. An alternative way to think about the improvement threshold is in percentage terms of an 'improvement grade'. Suppose there are 10 items. The maximum improvement grade is 10; that is, they improve on all items. The minimum is 0. Requiring that an individual improve on at least one item is requiring that the individual gets an improvement grade of 10% or better. So an obvious question becomes, where should the improvement grade threshold be set? Is 10% too low? Is 90% too high? Stated alternatively, what constitutes a passing grade? As it turns out that the measure of improvements vary significantly with the choice of the thresholds. If the threshold is set too low (as is currently done by the USDA which sets 1 as improvement threshold for the *adult* EFENP using the 10 item behavioral checklist), the outcomes are probably greatly overestimated where almost all individuals who participate would improve.

There is still another caveat that follows due to the choice of threshold. It turns out that when the thresholds for improvements are set too low, the probability that any individual improves will converge to one irrespective of the intervention. To demonstrate this claim, let us examine the probability of improvements, by employing a test consisting of "n" items, at two times. Assuming the response to each questions used to measure the outcomes as independent

trails, the probability of any outcome in n independent trials can be derived using the formula for the binomial probability distribution

$$P(Y = y) = \binom{n}{y} p^y q^{n-y}, \text{ where } \binom{n}{y} = \frac{n!}{(n-y)!y!}$$

where, p is the probability of success (or improvement); $q\ or\ (1-p)$ is the probability of failure (or no improvement); and Y is the outcome ($Y = 1$ if improve; $Y = 0$ if no improve). For n independent items, the values the outcome Y can take can be 0,…,n. The interest is in estimating $P(Y \geq 1)$ which is equivalent to the USDA measure of outcome (improvement in at least one question). We know that $P(Y \geq 1)$ is given by the sum of probabilities for all values of outcome greater than 1. Consider the case with 4 items ($n = 4$), then $P(Y \geq 1)$ will be equal to $P(Y = 2) + P(Y = 3) + P(Y = 4)$. That is, the probability of improvement is an increasing function of n (the number of questions under consideration). The probability of $P(Y \geq 1)$ converges to one as the number of questions under consideration increases. This result holds irrespective of whether the educational intervention is implemented or not. In other words, there is always a positive probability that the individual who takes the same test at any two time periods are likely to get the improvements by sheer chance. Even worse is that the probability actually converges to unity as the number of items gets large. The probability of improvement due to chance can exceed what is being measured as program outcomes by the USDA when the threshold for improvement is set too low. This is a caveat of using such improvement measure and is not unique to this dissertation.

Given the measurement caveats, the results must be interpreted with some caution as what the measures are picking up may simply be an artifact of the measure used and not the true impacts of the programs. The issue of optimal threshold is thus critical to accurately portrait the

program effectiveness and we think that the USDA should pay a closer attention and sort deliberation for improving their current measure of EFNEP outcomes.

Let aside the measurement caveat, overall, this dissertation contributes to the effectiveness of the EFNEP in several ways. First, it quantifies the budget effects of the adult EFNEP at the national level. This quantification provides an objective valid effectiveness criterion which can be used as benchmark to evaluate the relative effectiveness of the program. Second, the dissertation develops a framework for conducting the cost effectiveness analysis for the youth EFNEP program. As a consequence, a standardized instrument for the evaluation of youth EFNEP outcomes is developed which is a contribution in itself amid the lack of evaluation instruments to measure youth EFNEP outcomes. The availability of the effects instrument would likely motivate future stream of studies related to cost and effectiveness of the youth EFNEP in other states. Third the attributes of effective program for both adult and youth programs has been identified. Equipped with the information on cost and effectiveness, the EFNEP administrators can utilize such information to identify the programs that are doing relatively better than the others and share the practices of the successful programs. Lastly, chapter 4 applies a Rasch-type measurement model to measure the outcomes of the youth EFNEP, a noble approach (at least with respect to measuring outcomes of EFNEP). The USDA should consider this approach to measure EFNEP outcomes which surmounts some of the limitations and caveats in the current practice of outcomes measurement.

References

Baral, R., G.C. Davis., S. Blake., W. You., and E. Serrano. 2013. Using National Data to Estimate Average Cost Effectiveness of EFNEP Outcomes by State/Territory. *Journal of Nutrition Education and Behavior,* 45(2): 183-187.