

Using a hybrid design to analyze effectiveness and implementation of a refined energy-balance education module for upper elementary physical education

Senlin Chen*, Yang Liu* & Gregory Welk**

*School of Kinesiology, Pedagogical Kinesiology Laboratory, Louisiana State University, Baton Rouge, LA, USA

**Department of Kinesiology, Physical Activity and Health Promotion Laboratory, Iowa State University, Ames, IA, USA

Abstract

Informed by a previous experiment, this study sought to examine the effectiveness and implementation of a refined energy-balance (EB) education module for upper elementary physical education (PE). The refined module was distributed through a participatory research network to six schools in Iowa (United States) participated in the study and the evaluation followed the type I effectiveness-implementation hybrid design. The PE teachers completed standardized log sheets after delivering each lesson. Students in fourth and fifth grades ($n = 570$ and $n = 586$ at baseline and post-test, respectively) completed a written test before and after the experiment, to assess EB knowledge. The overall completion rate of the log sheets was low and varied across lessons, but process data meaningfully informed the implementation. The EB knowledge gain over time was modest and statistically significant in both boys and girls in fourth and fifth grades. The experienced schools showed greater knowledge gains than the naive schools that had not previously been exposed to the EB module. The findings support the potential of the refined EB education module for broader dissemination in elementary schools.

Keywords: effectiveness, elementary school, energy balance, implementation, participatory research network.

Résumé

Basée sur une expérience précédente, cette étude visait à examiner l'efficacité et la mise en œuvre d'un module d'éducation affiné à l'équilibre énergétique en éducation physique (EP) à la fin de l'école élémentaire. Le modèle affiné a été mis en œuvre à travers un réseau de participation à la recherche et six écoles dans l'état de l'Iowa aux Etats-Unis ont participé à l'étude et à l'évaluation suivi d'une conception hybride d'efficacité-mise en œuvre de type

I. Les enseignants d'EP ont complété un fichier standardisé de données après chaque leçon enseignée. Des élèves de CM1 (n=570) et CM2 (n=586) ont passé un test écrit avant et après l'expérimentation afin d'évaluer leurs connaissances dans le domaine de l'équilibre énergétique. Le taux global de renseignement des fichiers était faible et variait à travers les leçons mais la collecte des données a documenté de manière significative la mise en œuvre. On observe que le gain de connaissances dans le domaine de l'équilibre énergétique est modeste à travers le temps et significativement différent parmi les garçons et les filles en CM1 et CM2. Les écoles impliquées dans l'expérimentation ont des gains de connaissances plus importants que les premières écoles qui avaient précédemment mis en œuvre le module d'éducation à l'équilibre énergétique. Les résultats soutiennent le potentiel du module affiné d'éducation à l'équilibre énergétique pour une dissémination plus large dans les écoles élémentaires.

Mots clés : efficacité, école élémentaire, équilibre énergétique, mise en œuvre, réseau de participation à la recherche.

1. Introduction

Teaching children the basic principles of energy balance (EB) is crucial for them to learn how to establish and maintain a healthy body weight over time (Chen & Nam, 2017; Manore et al., 2014; Pasco & Ennis, 2015). In the United States (U.S.), physical education (PE) is an important setting through which students can gain EB knowledge and establish lifestyle behaviors for long-term weight control (Bassett et al., 2013). Traditionally, EB concepts, principles, and strategies are rarely taught in the gymnasium during PE, although they do make their appearance in health education textbooks at the secondary school level (e.g., late middle school or early high school years) (Friedman, Stine, & Whalen, 2009). The fact that PE and health education are offered as two separate subjects and sometimes taught by two different teachers, makes it difficult for students to connect their experience with learning across classrooms, hence limiting the potential for students to construct a deeper understanding of the content (Chen & Chen, 2012, 2014). Formal health education in the U.S. public schooling usually does not start until middle school years, but it would be more effective if youth already had foundational knowledge and skills related to healthy-living. Formalized EB education in upper elementary grades is particularly important to counter the high prevalence of childhood obesity rate and declines in physical activity that tend to occur through adolescence. Previous research has shown that EB education in upper elementary

school PE programming is both feasible and effective (Chen, Zhu, Androzzi, & Nam, 2018). However, the previous study only demonstrated the educational module's efficacy or internal validity under idealized and relatively controlled circumstances. Thus, additional research is needed to test effectiveness in broader and less controlled settings.

The dissemination and implementation of evidence-based curricula is critical to strengthen the quality of educational programming in PE (Giblin, Collins, MacNamara, & Kiely, 2014). Some exemplary curricula such as the *Science, PE, and Me* curriculum and the *Fitness for Life* have demonstrated success in enhancing student learning and are adopted by PE teachers for implementation in numerous schools (Chen, Martin, Sun, & Ennis, 2007; Sun, Chen, Zhu, & Ennis, 2012; Thompson & Hannon, 2012). However, many curriculum intervention trials stop at end of the efficacy-testing phase. A curriculum with established efficacy (using an exploratory trial) may have shown adequate internal validity to produce expected educational outcomes under ideal and controlled circumstances; however, it cannot be assumed to work when applied in less structured or un-controlled settings (Godwin et al., 2003; Halbert & MacPhail, 2010). The under-emphasis on external validity of interventions does not just exist in PE pedagogy research, but it is also a pressing concern in other fields such as public health or medicine, in which there are only a limited number of interventions that have studied effectiveness (pragmatic trial) and sustainability (Colditz, 2012; Green & Nasser, 2012). A recent review found that only a small percentage of the health interventions have systematically examined effectiveness, adoption, implementation, and maintenance, which has slowed down the speed of their translation to dissemination for large practical impact (Green & Nasser, 2012). Additional research examining the dissemination and implementation of interventions is critical to promote the translation of science into practice (Green & Nasser, 2012; Welk, 2017).

1.1. The effectiveness-implementation hybrid design

One solution to address the existing concerns about external validity of interventions is to study effectiveness and implementation simultaneously using a hybrid design (Curran, Bauer, Mittman, Pyne, & Stetler, 2012). Concerned about the slow translation of health promotion science, Geoffrey Curran and colleagues critiqued the issues inherent in the traditional step-wise clinical trials (from efficacy and effectiveness to dissemination) and subsequently proposed the *effectiveness-implementation hybrid design* (Brown et al., 2017; Curran et al., 2012). The hybrid design is positioned to expedite the process needed for science-informed or evidence-based practice and three types of hybrid designs were proposed (Curran et al., 2012). The type I hybrid design is focused on testing effects of a

health promotion intervention on outcomes (primary purpose) while observing and gathering information on implementation (secondary purpose) (Curran et al., 2012; Wolfenden, Williams, Wiggers, Nathan, & Yoong, 2016). Type II hybrid design places equal emphases on determining the intervention effectiveness as well as implementation strategy; while type III hybrid design primarily focuses on determining the utility of an implementation strategy but also secondarily addresses intervention outcomes (Curran et al., 2012). Following the type I hybrid design, we distributed and evaluated the refined EB education module through a participatory research network (PRN) called the called the *Iowa FitnessGram Initiative* (<http://www.iowafitnessgram.org>).

1.2. Description of the Iowa FitnessGram Initiative

The Iowa FitnessGram Initiative (<http://www.iowafitnessgram.org>) is a PNR of K-12 schools / teachers and university researchers committed to promoting youth physical fitness and physical activity in schools. As a PRN, it is designed to provide mutual benefits through participation, training, communication, implementation, and evaluation. The advantage to schools is that the network provides resources and training aimed at enhancing the assessment and promotion of physical activity and physical fitness (e.g., FitnessGram assessment in-service training; provision of programming). The advantage to the research group is the ability to deploy training and programs to learn how they work under real-world conditions. The prominent NFL PLAY 60 FitnessGram Partnership project was established as a distributed PNR (Welk, Bai, Saint-Maurice, Allums-Featherston, & Candelaria, 2016) and provided a model for the more localized Iowa FitnessGram Initiative. The goal was to promote bilateral communications between researchers and practitioners for sustained collaborations, leading to a harmonious integration of science and practice (Welk, 2017). It proved to be a particularly useful setting to distribute the refined EB education module for evaluation in this study.

1.3. Description of the EB Education Module

The EB Education module was designed and validated by a panel of elementary PE teachers, teacher educators, and researchers to teach students EB concepts, principles, and strategies through movement tasks in upper elementary school grades (Chen et al., 2018). The lessons are aligned with the U.S. national PE standards (Society for Health and Physical Educators, 2014) and use developmentally appropriate tasks to teach boys and girls in fourth and fifth grades relevant content in the psychomotor, cognitive, and affective domains. Each lesson is centered on a focal daily key question guiding students' learning of

EB concepts. For example, one lesson that addresses physical activity intensity has a focal question stated as: *How do you know if your heart is working hard enough to improve fitness?* (Answer: Take your pulse to see if you are in the Target Heart Rate Zone.) In this lesson, students learn about maximal heart rate (220-age) and Target Heart Rate Zone (50-85% of maximum heart rate) through movement stations categorized by light, moderate, or vigorous intensity physical activities. The stations allow them to learn how to quickly and accurately take pulses (carotid and radial pulses) and how to monitor and regulate exercise intensity. At end of the lesson, they are expected to name and differentiate activities of varying intensity categories. The optional incorporation of pedometers also enables students to better understand exercise intensity (steps per minute) and links to calories. This is just one example but all lessons follow a lesson plan script that unfolds in three sequential parts: a 5-minute teacher-selected warmup, a 20-minute “think as they move” main activity, and a 5-minute closure. For each one, the students are instructed to complete a task sheet in each lesson, to make connections between physical and cognitive engagements within the lessons.

Our previous experiment established the efficacy of the EB education module in four elementary schools (randomized to experimental or control schools) (Chen et al., 2018). The PE teachers in the experimental schools were instructed to utilize the 11 lesson plans, while the PE teachers in control were asked to follow their regular curriculum (Chen & Chen, 2012). That study demonstrated that both boys and girls in fourth and fifth grades who received the EB education significantly increased their EB knowledge (by ~15% as measured by a valid knowledge test) while also maintaining a similar level of in-class physical activity (measured by Actigraph GT3X+ accelerometers) as the students in the control condition (Chen et al., 2018). During the implementation, the PE teachers closely adhered (~90%) to the scripted PE lesson plans (as assessed by field observation checklists) and they also provided useful feedback for future improvement and modifications (as assessed by teacher interviews) (Chen et al., 2018).

The present study used the refined versions of these 11 lessons along with 5 new ones. Four of the schools in the evaluation were involved in the previous project (called experienced schools) and two were new schools (called naïve schools). This follow-up study was essential to determine whether the lessons could be implemented under less controlled settings with more distributed training methods.

1.4. Research purpose

This study capitalized on testing the effectiveness and implementation of the refined EB education module when distributed through a PRN – the Iowa FitnessGram Initiative. Guided by the effectiveness-implementation hybrid design (Curran et al., 2012), the primary purpose of this study (*effectiveness* testing) was to examine the difference in EB knowledge between baseline and post-implementation of the refined module, and differences in knowledge gains between experienced schools and inexperienced schools and by grade (i.e., fourth grade and fifth grade) and gender (i.e., boys and girls). The secondary purpose of the study (*implementation* testing) was to capture the process of implementing the module through process evaluation.

2. Method

2.1. Setting and participants

Six elementary schools recruited from the state of Iowa (U.S.A.) participated in this study. Table 1 shows the sociodemographic characteristics of these schools. Four of the six schools with previous exposure to the curricula were in the same small town (i.e., returning schools) while the other two (i.e., new schools) in remote rural areas. The small town schools were about twice as large as the rural schools. The schools showed varying levels of free or reduced lunch eligibility percentage, ranging from 22.1% to 58.7% ($M/SD = 36.8\%/12.2\%$). The free or reduced lunch eligibility is a frequently used measure by education researchers to capture school-level socio-economic status in the U.S. public school system (Sirin, 2005). The sample involved 570 and 586 fourth and fifth grade students at baseline and post-test, respectively. A majority of the students enrolled in these schools were Caucasian/White, which is typical for ethnic composition in the state. The study was approved by the Iowa State University *Institutional Review Board* (IRB) and the participating school districts. The approved IRB protocol waived collection of parental consent and assent because the study was rated as minimal risk and that de-identified data were collected.

Table I : Characteristics of the participating schools (n = 6)

School	Locale	School Enrollment	4 th Grade Enrollment	5 th Grade Enrollment	Student/Teacher Ratio	Free/Reduced Lunch Eligibility	Ethnicity White/Total [%]
1	Town	415	64	83	13.92	163/415	359/415
	Fringe					(39.3%)	(86.5%)
2	Town	402	68	71	15.03	89/402	379/402
	Fringe					(22.1%)	(94.3%)
3	Town	412	50	45	18.74	127/412	381/412
	Fringe					(30.8%)	(92.5%)
4	Town	453	65	92	14.98	153/453	417/453
	Fringe					(33.8%)	(92.1%)
5	Rural	246	30	32	16.51	88/246	239/246
	Remote					(35.8%)	(97.2%)
6	Rural	213	31	32	11.95	125/213	189/213
	Remote					(58.7%)	(88.7%)

2.2. Research design and procedures

This study employed the one-group pretest-posttest research design (Thomas, Nelson, & Silverman, 2015). Unlike the previous efficacy study (Chen et al., 2018), the six PE teachers (one per school) in this study received minimal training because the focus was to examine the natural adoption and implementation of the refined EB education module. Two weeks prior to the start of implementation (February 2016), the schools received the teacher's manual and ancillary materials (e.g., pedometers, task sheets, PE posters). One week before starting, students in fourth and fifth grades completed the baseline assessments. Then, the implementation started and lasted for 12 consecutive weeks (excluding the spring break week in March). In the course of implementation, the research team conducted three checkpoint sessions with the school implementation teams to monitor the process and provide guidance to schools. The checkpoints (~20mins per session) involved telephone conversations between the researchers and the school implementation staff at approximately weeks 4, 8, and 12. In addition, the PE teachers received weekly e-Newsletters through email that included teaching tips and resources to facilitate implementation. The PE teachers completed standardized log sheets to document the

process of implementation. Lastly, the students completed the same assessments at end of the implementation in May.

2.3. The Refined EB Education Module and its distribution

The refined EB education module consisted of 16 scripted lessons after adding five new lessons to the original 11-lesson version. These added content on sedentary behavior and screen time mitigation, in addition to physical activity and healthy-eating promotion (i.e., consuming more fruit and vegetables). Table 2 below shows the scope and sequence of the refined module. In this present study, PE teachers were advised to teach one PE lesson per week and were afforded autonomy to choose 12 of the 16 lessons for implementation. Furthermore, at their discretion, the PE teachers were encouraged to make and document modifications to the lessons to fit local characteristics and meet school needs (modifications documented on log sheets). The rationale for making the three above arrangements was to empower the PE teachers with choices and autonomy to result in more motivated adoption and implementation of the refined module.

Table II : The Scope and Sequence of SWITCH PE (16 Lessons)

Lesson	Concepts	Main Activity
1	Physical activity; intensity; steps measurement	“Count My Steps”
2	Screen time; commercial break physical activity	“Switch Your VIEW – DAY 1”
3	Food groups; energy in	“Choose My CHEWs”
4	Energy in; energy balance; fruits and vegetables	“Capture the Fruits and Veggies”
5	Physical activity intensity; target heart rate zone	“Target Heart Rate Zone”
6	Screen time; commercial break physical activity	“Switch Your VIEW – DAY 2”
7	Food groups; energy in; balanced meal; empty calories	“Snack Attack”
8	Energy in; energy balance	“Bowl to Balance”
9	Energy out; physical activity, intensity	“Energize My Steps”
10	Screen time; commercial break physical activity; technology	“Switch Your VIEW – DAY 3”
11	Energy in/balance; physical activity	“Energy Beanbags”
12	Fat; energy balance	“Healthy and Unhealthy Fat”
13	F.I.T.T. principles; energy out; energy balance	“Have a F.I.T.T. Body”

14	Screen time; commercial break physical activity; technology	“Switch Your VIEW – DAY 4”
15	Food groups; physical activity; energy balance	“Eat to Move”
16	Body composition; fat tissues; lean tissues; healthy diet; exercise	“Fat Cell Tag and Ultimate CHEW”

2.4. Variables and instrumentation

PE teachers' implementation of the refined EB education module was assessed using log sheets. The log sheets were developed by our research team to monitor the extent to which the teachers would follow the lesson plans. Each log sheet had 10 questions asking teachers to document how they implemented each lesson. Below shows some example questions. Nine of the 10 questions were close-ended and categorical (i.e., multiple choices or Yes/No), while the final question asked for short answers. Below shows the four sample questions.

#2. *How long did the lesson last?* [5min, 10min, 15min, 20min, 25min, or 30min]

#3. *Did you modify the activity* [No, Yes; if yes, how? _____]

#7. *How physically engaged were the students during the PE lesson?* [1. Not at all, 2. A little, 3. Very]

#9. *Did the students enjoy the lessons?* [1. Not at all, 2. A little, 3. Very]

#10. *Did you face any challenges or are there any comments you would like to make?*
[short answer]

We used a previously validated *Energy Balance Knowledge Test* (Chen, Zhu, & Kang, 2017) to measure EB knowledge performance in students. Due to the added focus on sedentary behavior or screen time in the lesson plans, we added four questions to further assess students' knowledge in this behavior domain. Therefore, the current version of the knowledge test consisted of 28 multiple-choice questions (with four possible answers), capturing knowledge across three behavior domains (Do = physical activity, View = sedentary behavior or screen time, and Chew = fruits and vegetable consumption) and the overall EB. These questions test students' understanding of EB-related concepts including *energy balance, energy intake, energy expenditure, as well as sub-concepts such as macronutrients, Calorie (kcal), food groups, physical activity, intensity, steps, body composition, screen time, etc.* (see Table 2 for details). For example, a question measuring knowledge about body composition is phrased as: *Which term relates to the amount of muscle, bone, and fat you have in your body?* a. Aerobic endurance, b. Muscle endurance,

c. *Flexibility*, and d. *Body composition* (correct answer). The original test has shown very good validity based on the Rasch Model analysis in our previous study (Chen et al., 2017).

2.5. Data collection and reduction

We collected the completed log sheets from the schools and then a trained data analyst entered the data into a data processing computer. We distributed the EB knowledge test through a www.Qualtrics.com survey to the schools where students responded in their media centers with technical assistance from their PE teachers. During the testing process, we instructed the students to try their best to answer the questions and that their responses would not affect their grades or school standing. After data collection, we scored the test sheets in accordance to the answer key (1 = correct; 0 = incorrect). EB knowledge performance was determined by computing the percentage of correct scores (i.e., sum of correctly answered questions divided by 28 questions).

2.6. Data analysis

We analyzed teacher' log data using descriptive analyses of the individual items (i.e., Mean, Standard Deviation, and Percentage) to quantify the EB education module implementation. We next analyzed the differences in EB knowledge by time (pretest vs posttest), group (4 experienced schools vs. 2 naïve schools), grade (fourth vs fifth grades), and gender (boys vs girls) using analysis of variance (ANOVA). To further depict the change patterns of the overall EB knowledge scores, we conducted cluster analysis to classify scores into three knowledge levels (i.e., high, moderate, and low) followed by ANOVA to verify the group classification. Based on the classifications, we evaluated the migration of scores across knowledge level categories between pretest and posttest (e.g., fewer low knowledge students and more high knowledge students would suggest knowledge gain) using descriptive statistics and Pearson's Chi-square analysis. We also compared knowledge gain scores in the three behavior areas related to EB (i.e., knowledge of *Do*, knowledge of *View*, and knowledge of *Chew*). Cohen's d was computed as effect size for the mean difference comparisons (small = 0.20, medium = 0.50, large = 0.80).

3. Results

3.1. PE teachers log sheets

Although we encouraged the PE teachers (one PE teacher per school) to complete the log sheets as soon as they finished teaching each lesson, we did not *require* them to fulfill this task as a requirement in this study. Only three PE teachers (out of six) from three different

schools (schools 2, 5, and 6) completed the log sheets for a minimum of one lesson. The frequency of logs completed per PE lesson ranged from 0 (lessons 1, 6, and 7) to 8 (lesson 4), with median being 5. Descriptive analysis of these data showed that about half of the PE lessons were taught without modifications from the lesson plans (24/47, 51.2%), while some kind of modifications occurred with the other lessons. Most of the lessons lasted for 30 minutes (41/47, 87.2%) which was the standard duration length of PE classes offered in these six schools, with one lesson lasting 15 minutes, two lessons 20 minutes, and one lesson 25 minutes for uncertain reasons (e.g., early dismissal, incomplete implementation, etc.). The PE task sheets, as ancillary instructional materials, were under-utilized (18/47, 38.3%). The PE teachers indicated that they were mostly satisfied with the lessons (33/47, 71.2%) and the lessons had appropriate level of difficulty for the students (43/47, 91.5%). They also perceived that the students enjoyed the PE lessons (32/47, 89.3%) and showed a good level of engagement during the lessons (45/47, 95.8%).

3.2. EB knowledge gains

EB knowledge showed consistent gains across the six schools with an average 8.5% increase from pre-test to post-test (ΔM ranging from 2.1% to 12.8% across the schools; Cohen's $d = 0.56$). This increase was statistically significant ($F_{1,1155} = 91.98, p < .01, \eta_p^2 = .07$). Subsequent cluster analysis demonstrated that the percentage of students classified as high knowledge level increased from 20.9% in pre-test to 46.6% in post-test, while the percentage of students classified as low level decreased from 41.2% in pre-test to 19.8% in the post-test. Table 3 presents the knowledge groups classification thresholds and group distribution (% of students) as well as the Pearson's Chi-square test result ($\chi^2 = 91.75, p < .01$).

Table III : Students Classified as High, Moderate and Low Knowledge Levels at Pre-Test (n = 570) and Post-Tests (n = 586)

Knowledge Levels	Classification Threshold (Performance Scores)		Learning Result (% of Students)		Pearson χ^2 (df = 2)
	Min	Max	Pretest	Posttest	
High knowledge	66.67%	96.30%	20.88%	46.59%	91.75**
Moderate knowledge	51.85%	62.96%	41.93%	33.62%	
Low knowledge	14.82%	48.15%	37.19%	19.80%	

Note. Min = minimum; Max = maximum. ** $p < .01$

Knowledge gain was larger in the four experienced schools ($\Delta M = 9.6\%$, Cohen's $d = 0.72$) than the two naïve schools ($\Delta M = 2.8\%$, Cohen's $d = 0.21$). In addition, knowledge gains were similar between boys ($\Delta M = 8.1\%$, Cohen's $d = 0.51$) and girls ($\Delta M = 8.8\%$, Cohen's $d = 0.62$), and between fourth grade ($\Delta M = 9.7\%$, Cohen's $d = 0.73$) and fifth grade classes ($\Delta M = 6.9\%$, Cohen's $d = 0.45$).

To determine knowledge gain in the three behavior domains related to EB, we computed the mean differences in knowledge of Do (physical activity), knowledge of View (sedentary behavior and screen time), and knowledge of Chew (dietary behavior) between pre-test and post-test. Figure 1 illustrates these changes. Knowledge gains were observed in all three behavior domains (knowledge of Do: 12.57%, knowledge of View: 6.06%, knowledge of Chew: 5.08%) meaning the students enhanced their knowledge in all three knowledge areas.

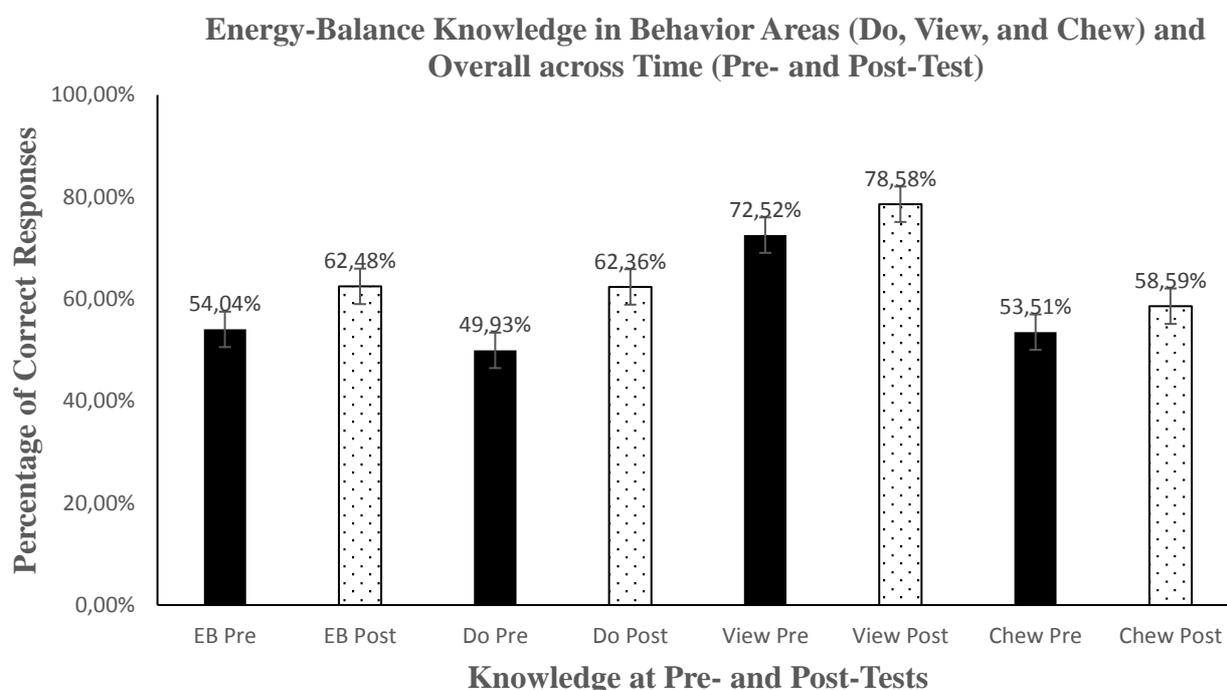


Figure 1 : Energy-Balance Knowledge Performance Before ($n = 570$) and After ($n = 586$) Module Implementation

4. Discussion

The results from this study support the effectiveness and implementation of the refined EB education module for upper elementary school PE, when distributed through a PRN called the Iowa FitnessGram Initiative. Boys and girls in fourth and fifth grades across the six schools showed significant gains of EB knowledge (with moderate effect sizes), although

the students from the experienced schools showed greater gains than those from naive schools.

The main goal of the refined EB education module was to promote knowledge about EB principles through active tasks during PE classes. The lessons afforded students the opportunity to engage in the carefully designed learning tasks (i.e., “think as they move”) in order to acquire an enhanced understanding of EB and its related behaviors. The original module demonstrated efficacy in promoting EB through active PE lessons (Chen et al., 2018), so we evaluated the effectiveness of the refined module using a type I *effectiveness-hybrid design* (Curran et al., 2012). Specifically, the participating schools implemented the refined module with little training and support received, through the robust PRN (i.e., Iowa FitnessGram Initiative). The goal of releasing them in this way was to assess whether the lessons could be distributed this way to promote broader dissemination. Indeed, our results confirmed a significant gain of EB knowledge ($\Delta M = 8.5\%$, ranging from 2.1% to 12.8%, $d = 0.56$) among the students, although the degree of change was smaller than that observed in the previous efficacy study ($\Delta M = \sim 15\%$ increase; Chen et al., 2018). The EB knowledge gain was also evidence among both boys and girls in fourth and fifth grades (Cohen’s d ranged from 0.45 to 0.73). These significant knowledge gains were comparatively smaller than those found in larger-scale multi-year PE-based curriculum interventions such as the Science, PE & Me curriculum (Cohen’s d ranged from .97 to 2.21) (Sun et al., 2012), but our study involved significantly lower intervention dosage and was distributed by the Iowa FitnessGram Initiative PRN in the naturalistic and less controlled settings.

Students’ learning improvement was also supported by the migration pattern of dwindling student counts classified in the low knowledge group versus the increasing student counts classified in the high knowledge group, as shown in our cluster analysis. Knowledge gains were also evident in the three individual behavior domains, with the largest gains in knowledge of physical activity ($\Delta M = 12\%$) followed by knowledge of sedentary behavior ($\Delta M = 6\%$) and knowledge of eating behavior ($\Delta M = 5\%$). Students from all six schools, regardless of their grade and gender, all showed the same increasing pattern for EB knowledge, albeit the more experienced schools showed greater gains than the naïve schools. These findings support the effectiveness of the refined EB education module, when distributed through the Iowa FitnessGram Initiative. The Iowa FitnessGram Initiative provided a useful structure to distribute, facilitate, and evaluate the use of the modules under real-world conditions.

The overall smaller knowledge gains observed in this effectiveness study compared to the previous efficacy study (Chen et al., 2018), as well as the smaller gains observed in the naïve schools compared to the experienced schools, suggest some loss of fidelity when less structure is provided. The process data showed a low completion rate of the log sheets, which hampered our ability to fully understand the teachers' implementation process across the six schools. Future research should use multiple forms of quality process measures (e.g., checklists, interviews) to keep the implementation process accountable and informative, as teachers proceed with intervention. Asking teachers to voluntarily report the degree of implementation is relevant to researchers but may not be perceived as relevant and important to practitioners. Therefore, future studies should identify strategies to improve completion of self-reported fidelity data in a more consistent manner. Recognitions, incentives, reminders, and encouragements could be provided through the PRN so the participating teachers will see process evaluation as an important step of program evaluation. Nonetheless, in spite of not being able to see the full implementation landscape, we did gather some encouraging and useful insight about the module implementation. From the completed log sheets, we learned that the refined EB education lessons led to overall positive teaching (high teacher satisfaction, appropriate content difficulty, with modest level of lesson modifications) and learning (enjoyable experiences and modest engagement). From these process data, future adaptations can be made to facilitate easier access and implementation. For example, we believe that communication with the PE teachers may be more effective when coordinated messages are coming from within the school system (i.e. from another member of the school such as principal or colleague). We will also attempt to furnish lesson plans and ancillary instructional materials (i.e., task sheets) in simpler and shorter forms. The broader release of the materials through the Iowa FitnessGram Initiative and other PE related networks would enable teachers to adopt and adapt these materials to fit their own needs.

A major strength of the study lies in that we followed Curran et al.'s effectiveness-implementation hybrid design (type I) to evaluate both the effectiveness and implementation of the refined EB education module. This research investigation highlights the importance of studying the external validity of curriculum interventions in PE pedagogy research, by following a robust hybrid research design. The findings confirm the utility of the Iowa FitnessGram Initiative as a PRN for program evaluation. The other strength of the study is related to the continuity of our research on EB education through school PE. The findings from our previous efficacy study and this effectiveness study provide a more complete story

on steps to translate an evidence-based curriculum or program into a practical set of instructional strategies.

Despite the aforementioned strengths, we also acknowledge several limitations. First, although the primary purpose of this study was to evaluate the effectiveness of the refined EB education module, we had significant missing data (self-reported) to fully address the secondary purpose (implementation testing). The completed log sheets (with low completion rate) were the only process data used to describe the implementation process, limiting our ability to observe the full picture of implementation as well as the potential moderating effect of implementation on intervention effectiveness. The next step of research effort would be to carry out Curran et al.'s type II hybrid research design that treats effectiveness and implementation testing as two equally important aims (Curran et al., 2012). The other potential limitation of the study is the lack of a control group for effectiveness testing. Traditional effectiveness studies often involve a randomly assigned control group. Having a control group would definitely strengthen the evidence of our effectiveness trial, but it would also increase the difficulty for feasibility and practicability. In addition, our previous efficacy study already showed significant outcomes when compared to control schools, thus we did not see having control schools extremely important in this round of evaluation. Our cross-sectional comparisons of knowledge gains between school types (experienced vs. naïve), grade (fourth vs. fifth grades), and gender (boys vs. girls) along with process evaluation data in this study, provide rich details about both the module effectiveness and implementation.

References

- Bassett, D. R., Fitzhugh, E. C., Heath, G. W., Erwin, P. C., Frederick, G. M., Wolff, D. L., . . . Stout, A. B. (2013). Estimated energy expenditures for school-based policies and active living. *American Journal of Preventive Medicine, 44*(2), 108-113. doi:10.1016/j.amepre.2012.10.017
- Brown, C. H., Curran, G., Palinkas, L. A., Aarons, G. A., Wells, K. B., Jones, L., . . . Cruden, G. (2017). An overview of research and evaluation designs for dissemination and implementation. *Annual Review of Public Health, 38*, 1-22. doi:10.1146/annurev-publhealth-031816-044215
- Chen, A., Martin, R., Sun, H., & Ennis, C. D. (2007). Is in-class physical activity at risk in constructivist physical education?. *Research Quarterly for Exercise and Sport, 78*, 500-509.

- Chen, S., & Chen, A. (2012). Ninth graders' energy balance knowledge and physical activity behavior: An expectancy-value perspective. *Journal of Teaching in Physical Education*, 31, 293-310.
- Chen, S., & Chen, A. (2014). Ninth graders' learning differences in a healthful-living curriculum. *Learning and Individual Differences*, 30, 170-176. doi:10.1016/j.lindif.2013.11.009
- Chen, S., & Nam, Y. H. (2017). Energy balance education in schools: The role of student knowledge. *European Physical Education Review*, 23(2), 137-170.
- Chen, S., Zhu, X., Androzzi, J., & Nam, Y. H. (2018). Evaluation of a concept-based physical education unit for energy balance education. *Journal of Sport and Health Science*, 7(3), 353-362.
- Chen, S., Zhu, X., & Kang, M. (2017). Development and validation of an energy-balance knowledge test for fourth- and fifth-grade students. *Journal of Sports Sciences*, 35(10), 1004-1011. doi:10.1080/02640414.2016.1208837
- Colditz, G. A. (2012). The promise and challenges of dissemination and implementation research. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health* (pp. 3-22). New York: Oxford University Press.
- Curran, G. M., Bauer, M., Mittman, B., Pyne, J. M., & Stetler, C. (2012). Effectiveness-implementation hybrid designs: Combining elements of clinical effectiveness and implementation research to enhance public health impact. *Medical Care*, 50(3), 217-226. doi:10.1097/MLR.0b013e3182408812
- Friedman, D. P., Stine, C. C., & Whalen, S. (2009). *Holt lifetime health*. New York: Holt, Rinehart and Winston.
- Giblin, S., Collins, D., MacNamara, A., & Kiely, J. (2014). "Deliberate preparation" as an evidence-based focus for primary physical education. *Quest*, 66(4), 385-395.
- Godwin, M., Ruhland, L., Casson, I., MacDonald, S. M., Delva, D., Birtwhistle, R., . . . Seguin, R. (2003). Pragmatic controlled clinical trials in primary care: The struggle between external and internal validity. *BMC Medical Research Methodology*, 3, 38. doi:10.1186/1471-2288-3-28
- Green, L. W., & Nasser, M. (2012). Furthering dissemination and implementation research: The need for more attention to external validity. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health*. New York: Oxford University Press.

- Halbert, J., & MacPhail, A. (2010). Curriculum dissemination and implementation in Ireland: Principal and teacher insight. *Irish Educational Studies*, 29(1), 25-40. doi:10.1080/03323310903522677
- Manore, M. M., Brown, K., Houtkooper, L., Jakicic, J., Peters, J. C., Edge, M. S., . . . Krautheim, A. M. (2014). Energy balance at a crossroads: Translating the science into action. *Medicine and Science in Sports and Exercise*, 46(7), 1466-1473. doi:10.1249/Mss.0000000000000318
- Pasco, D., & Ennis, C. D. (2015). Third-grade students' mental models of energy expenditure during exercise. *Physical Education and Sport Pedagogy*, 20(2), 131-143.
- Society for Health and Physical Educators. (2014). *National standards & grade-level outcomes for K-12 physical education*. Urbana-Champaign, IL: Human Kinetics.
- Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.
- Sun, H., Chen, A., Zhu, X., & Ennis, C. D. (2012). Learning science-based fitness knowledge in constructivist physical education. *Elementary School Journal*, 113(2), 215-229. doi:10.1086/667405
- Thomas, J. R., Nelson, J. K., & Silverman, S. J. (2015). *Research methods in physical activity*. Champaign, IL: Human Kinetics.
- Thompson, A., & Hannon, J. C. (2012). Health-related fitness knowledge and physical activity of high school students. *Physical Educator*, 69, 71-88.
- Welk, G. J. (2017). The intersections of science and practice: Examples from FitnessGram (R) Programming. *Research Quarterly for Exercise and Sport*, 88(4), 391-400. doi:10.1080/02701367.2017.1377485
- Welk, G. J., Bai, Y., Saint-Maurice, P. F., Allums-Featherston, K., & Candelaria, N. (2016). Design and evaluation of the NFL PLAY 60 FITNESSGRAM partnership project. *Research Quarterly for Exercise and Sport*, 87(1), 1-13. doi:10.1080/02701367.2015.1127126
- Wolfenden, L., Williams, C. M., Wiggers, J., Nathan, N., & Yoong, S. L. (2016). Improving the translation of health promotion interventions using effectiveness-implementation hybrid designs in program evaluations. *Health Promotion Journal of Australia*, 27(3), 204-207. doi:10.1071/HE16056