HHS Public Access
Author manuscript
Contemp Clin Trials. Author manuscript; available in PMC 2018 October 01.

Published in final edited form as:
Contemp Clin Trials. 2017 October ; 61: 81–86. doi:10.1016/j.cct.2017.07.023.

A Cluster Randomized Control Trial to Assess the Impact of Active Learning on Child Activity, Attention Control, and Academic Outcomes: The Texas I-CAN Trial

John B Bartholomew[1], **Esbelle M Jowers**[1], **Vanessa L Errisuriz**[1], **Sharon Vaughn**[2], and **Gregory Roberts**[3]

[1]The Department of Kinesiology and Health Education, The University of Texas at Austin, Austin, Texas

[2]Department of Special Education, The University of Texas at Austin

[3]The Meadows Center for Preventing Educational Risk, The University of Texas at Austin

Abstract

Background—Active learning is designed to pair physical activity with the teaching of academic content. This has been shown to be a successful strategy to increase physical activity and improve academic performance. The existing designs have confounded academic lessons with physical activity. As a result, it is impossible to determine if the subsequent improvement in academic performance is due to: (1) physical activity, (2) the academic content of the active learning, or (3) the combination of academic material taught through physical activity.

Methods / Design—The Texas I-CAN project is a 3-arm, cluster randomized control trial in which 28 elementary schools were assigned to either control, math intervention, or spelling intervention. As a result, each intervention condition serves as an unrelated content control for the other arm of the trial, allowing the impact of physical activity to be separated from the content.

Address correspondence to: John B. Bartholomew, Department of Kinesiology and Health Education, The University of Texas at Austin, 2109 San Jacinto Blvd, Mail Stop D3700, Austin, TX, 78712-1204, [jbart@austin.utexas.edu], 512-232-6021.

Declarations:

Ethics approval and consent to participate. This study received ethical approval from the Institutional Review Board for Human Subjects Research (2011-01-0014) at The University of Texas at Austin. In addition, all procedures were reviewed and approved by each participating school district and school principal. Both parental consent and child assent were collected from all children participating in this study. Teachers also individually consented to participate in this study.

Consent for publication. Not Applicable

Availability of data and material. Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Competing interests: The authors declare that they have no competing interests.

Authors' contributions. JBB conceptualized and designed the study and drafted the initial manuscript. EMJ contributed to the conceptualization and design of the study, and reviewed and revised the manuscript. VLE coordinated and participated in data collection at all sites and reviewed and revised the initial manuscript. GR conceptualized analyses, conducted power analysis, and reviewed and revised the manuscript. SV conceptualized analytic strategy and reviewed and revised the manuscript. All authors read and approved the final manuscript.

That is, schools that perform only active math lessons provide a content control for the spelling schools on spelling outcomes. This also calculated direct observations of attention and behavior control following periods of active learning.

Discussion—This design is unique in its ability to separate the impact of physical activity, in general, from the combination of physical activity and specific academic content. This, in combination with the ability to examine both proximal and distal outcomes along with measures of time on task will do much to guide the design of future, school-based interventions.

Trial Registration—NCT03087279 (Retrospectively Registered 03/21/2017)

Keywords

children; active learning; elementary school; academic achievement

BACKGROUND

While childhood physical activity (PA) tracks into adulthood [1, 2], PA generally declines across childhood [3], with the greatest declines occurring during elementary school [4]. In addition, children spend up to 30 hours at school - 73% of that time sedentary [5]. Interventions to increase PA in elementary years is paramount. Moreover, interventions that compel activity during non-discretionary, school hours is especially likely to succeed, as these interventions consistently increase PA [6–8], and reach populations (e.g. high BMI, low SES, minority) most at-risk for sedentary lifestyle [15]. Physically active, academic lessons are one option as they compel PA by utilizing movement in the teaching of academic content [8–10]. For example, a one-year intervention with children 9–11 years found a 20% increase in cardiorespiratory fitness and a reduction in systolic BP [11], yet no change in BMI nor standardized testing [12]. In contrast, a two-year intervention found reductions in BMI and significant improvement on both math and spelling performance [6]. A two-year, cluster RCT, [13] found the intervention to improve math and spelling scores relative to controls; while another cluster RCT [14] only found improvements in math among children who performed most poorly at baseline.

Limitations of Previous Research

The present studies make it impossible to determine if improvement in academic performance is due to PA, academic content, or their combination. Additionally, we do not know how active learning impacts more immediate academic outcomes or mediators for academic improvement. One study [10] found that 10-minutes of active learning improved student focus, or time-on-task (TOT) relative to students in sedentary lessons whose TOT dropped significantly. We propose that brief periods of PA improve TOT independent of content and that this change – carried out multiple times over the school year – impacts short and long-term academic achievement.

Texas I-CAN Project

The Texas I-CAN (Initiatives for Child Activity and Nutrition) project is designed to address these limitations. This is a 3-arm cluster RCT: control schools; physically active math schools; or physically active spelling schools. All schools will be assessed for:

accelerometer-based PA; TOT; short-term (1–2 week) teacher made math and spelling tests; and long-term (9 months) standardized math and spelling tests. We will then examine PA, TOT and one-week academic performance as predictors of standardized test scores. Additionally, each intervention condition serves as an unrelated content / active control for the other. If both schools improve in math or spelling, this will indicate a general effect of PA. If improvement is limited to the content of their assigned arms (e.g. only math schools improve in math), this indicate an effect for the combination of PA and lesson content.

Aims

The **primary aim** is to evaluate effects of active learning on PA, TOT, and short and long-term academic performance in 4^{th} grade children. **Secondary aims** will: (a) test the mediating role of PA and TOT on academic outcomes; and (b) test the moderating role of fitness, SES, ethnicity, sex & BMI.

METHOD

Participants, ethics approval and consent to participate

All procedures were approved by the Institutional Review Board for Human Subjects Research at The University of Texas at Austin. In addition, all procedures were reviewed and approved by each participating school district and school principal. Once approved, fourth grade teachers and students were targeted for recruitment within each school. Teachers in schools were recruited as a team, and consent from each teacher was obtained for inclusion in the program. Twenty-eight schools (n=19 intervention and n=9 control) and 149 teachers (n=99 intervention, n=50 control) were recruited. Schools were, on average, 31.99% Hispanic, 9.51% Black, and 46.26% White. About 21% of students were eligible for free or reduced-priced lunch, and 28.5% classified as overweight/obese. In addition, parental consent and child assent were collected from 2,716 fourth grade children, making this the largest RCT to address this topic to date. Figure 1 shows the flow of schools and participants through the study.

I-CAN Teacher Training

Our implementation and training centered on the Theory of Planned Behavior [15], supplemented with key concepts from Social Cognitive Theory [16]. Social Cognitive Theory is based on a reciprocal determinism amongst a person, their environment and their behavior. As such, our training is designed to create a supportive environment in which teachers can develop the skills required to implement lessons through guided trials and group-based problem solving to overcome barriers. Specifically, the training was designed to have teachers implement all aspects of the intervention and evaluation process during a run-in period in the year prior to implementation. This allowed teachers to build self-efficacy and ensure the quality of their implementation. This training was supplemented with visits to each intervention school to review challenges and problem-solve solutions.

Procedures

Prior to and following the intervention, distal academic data related to math and spelling was collected from each school, with each group of schools assessed simultaneously. This was designed to ensure that any unforeseen environmental change (e.g. poor weather) was equally distributed across conditions. Additionally, student BMI and fitness data was obtained from school records. Preliminary teacher-level data was collected following the initial training/workshop and included: (a) the lesson rating scale; and (b) perceived behavioral control for lesson implementation.

Intervention data collection—includes data collected over one school week during the intervention. This included: (a) TOT observations; (b) in-school physical activity assessed through accelerometry; and (c) three sets of teacher-developed proximal outcomes (pre-test, posttest, retention) for spelling and mathematics. For intervention schools, TOT observations occurred prior to and following a Texas I-CAN active learning session. For control schools TOT observations occurred prior to and following the initial 15 min of their traditional lesson.

Measures

Demographic Data—The parental informed consent asked for release of demographic data from school records. These records were utilized to indicate sex, age and race/ethnicity. In addition, status as qualifying for free and reduced cost lunch was used as a proxy for a dichotomous indicator of SES. As an intervention of this length is unlikely to impact fitness or BMI, these were collected as demographic and grouping data. Participant BMI and fitness were derived from school FITNESSGRAM® [17] records. The FITNESSGRAM® includes tests of each component of physical fitness. Cardiorespiratory / aerobic fitness is assessed through the PACER, a 20-m shuttle run that progressively increases in difficulty. Muscular endurance and power are measured through push-ups and curl-ups. Flexibility is assessed through a sit and reach test. FITNESSGRAM® software provides an overall score and a fitness profile for each child and is widely considered a valid and reliable measure of physical fitness [17, 18] that is used in schools across the country. Each of our target school districts required a FITNESSGRAM® to be completed by all students in 3rd, 4th and 5th grade. The FITNESSGRAM® also includes measures of height and weight completed by the school nurse and PE teacher. These were used to calculate BMI.

Class-level data—Class-level data was collected from teachers as a part of our process evaluation that reflects the primary components of the TPB. Teacher attitude was assessed with 6 items (Cronbach's $\alpha = 0.91$) that ask teachers to rate how they feel about implementing I-CAN for 15 minutes a day, four days a week. Items are bipolar adjective pairs (e.g. unpleasant – pleasant; useless – useful; worthless – valuable) and were scored on a 7-point, Likert scale (from 1 to 7) and summed. Perceived behavioral control (PBC) was assessed with 12 items (Cronbach's $\alpha = 0.92$) that ask teachers whether they are confident they can implement I-CAN in the face of a series of constraining factors (e.g. preparing for standardized tests, inclement weather, etc.) along with their confidence that a series of facilitating factors will be present (e.g. easily incorporated into curriculum, easily modified to fit with curriculum, etc.). Items were scored on a 7-point Likert scale from 1 (Strongly Disagree) to 7 (Strongly Agree) and summed. This strategy will allow for a direct assessment of PBC, in line with the recommendations of [19].

Implementation Rates—As a process evaluation, we tracked implementation through: (1) teacher self-report; and (2) observations of lesson plans for each teacher. The observations were completed by trained research staff. Specifically, staff rated students' physical activity during I-CAN lessons. Ratings centered on how many children were active [(1) less than half of the class to (3) more than half of the class], how often children were active [(1) not at all to (5) most of the time], the intensity of movement for the lesson [(1) standing still to (5) running]. Observers' rating will be averaged across each variable and used in subsequent analyses. ICCs were calculated to determine IRR among observers for physical activity variables. ICCs were 0.93 for how often children were active and intensity of movement variables, and the ICC for how many children were active was 0.94, indicating excellent inter-rater reliability for all physical activity variables.

Outcome Measures

Physical Activity—Physical activity was assessed through minutes of moderate and vigorous physical activity (MVPA) via Actigraph accelerometry, which is widely considered the most valid and reliable accelerometer [20, 21]. It provides 14 days of continuous monitoring with 4MB of data storage. The GT3X+ was affixed above the iliac crest of the hip of each participant with an elastic belt within 30 min of the onset of school and was removed within 30 min of dismissal. This placement has been successfully used with children [4, 22]. Daily levels of MVPA vary widely for children, with few vigorous activities sustained for more than 10 minutes [4]. As a result, sampling epochs were set at 5 sec intervals to best capture variability in activity [23]. The GT3X+ was worn for a full school week - 5 consecutive days, from the onset to the end of the school day, which meets minimum recommendations for reliability [24, 25]. As the accumulated counts can be separated by time, these procedures allow for calculation of in-class physical activity, separating time in PE and recess. The resulting dependent variables are: (1) steps; and (2) time spent in moderate (4–5.9 METs) and vigorous (≥6 METs) physical activity.

Engagement with Academic Material – Time on Task—Direct observations of children using Momentary Time Sampling (MTS) was used to indicate time on task (TOT). In this procedure, children are observed for on/off task behavior. On-task behavior is defined as any behavior in which a student is attentive to the teacher or actively engaged in the assigned task. Off-task behavior is defined as behavior that does not fall under the specifications of "on-task" behavior. Examples of off-task behavior included a student: gazing off, placing his head on the desk, reading or writing inappropriate or unassigned material. TOT is then calculated for each student by dividing the number of on-task observations by the total number of observations per student. Because observations of more than 5 seconds were found to be less reliable [26], individual assessments were limited to a series of 5-second observations within a 15 min observation period, with two trained observers. Each observer recorded the behavior of one-half of the participants in each class. Fifteen minutes of observation provides 180, 5-second observations whereby each student was observed approximately 22 times/assessment. Observers were trained in pairs within a separate set of non-participating elementary classrooms to prevent contamination of the observations. Training centers on viewing, coding, and recording behavior in line with operational definitions of categories and discussion of examples within that class. Training is considered complete when inter-rater reliability exceeded 90%. In an earlier study [27], our research team found four training sessions were required to achieve sufficient inter-rater reliability of 92%. A three-month, follow-up assessment indicated that IRR remained high, at 94%.

Proximal (weekly) academic outcomes—The use of teacher-derived outcomes is problematic as they can differ greatly across teachers and schools. However, standardized testing fails to consider the specific content covered in a week-long set of lessons. In response, we utilized an independent teacher to develop a set of mathematics and spelling tests that reflect the school district requirements for 4th grade students. These were shared with teachers prior to implementation to ensure that their lessons during data collection week reflected this content.

Distal academic outcomes—The Gates-MacGinitie (Level 4) and specific subsections of the Woodcock Johnson III Normative Update Tests of Achievement (WJ-III ACH) were used to assess distal academic outcomes among participants. The achievement tests selected meet the highest psychometric standards for assessments of the skills needed to measure. The tests are demonstrated to be reliable and valid assessments. Use of these assessments for the I-CAN project allow us to easily compare our research results to other studies regarding student achievement.

The Comprehension section of the Gates-MacGinitie Reading Test Level 4 was given at baseline (Form S) and post-intervention (Form T). Each test consists of 11 short paragraph passages, equally divided between fiction and information texts, with a total 48 questions. Validity and reliability are known to be high [28]. For instance, the Kuder-Richardson 20 reliability with a fourth-grade sample is 0.93 [29].

The WJ-III ACH is a standardized, nationally norm-referenced achievement test and provides important data regarding student achievement in spelling and mathematics at pre- and post-test. The following subtests were used for this project: 1) Test of Math Calculation: 45 total problems (ranging from simple arithmetic to high school level calculations), timed test administered for 22.5 minutes (published median reliability for children is .85); 2) Test of Math Fluency (speed for simple addition, subtraction, and multiplication), timed test administered for 3 minutes (published median reliability for children is .89); and 3) Spelling Test, specifically questions 12 – 59, letters and words were read aloud by a proctor and spelled by students, allowing a maximum of 30 seconds for each item (test-retest reliability is .90 [30]).

ANALYSIS

Power Analysis

Sampling was organized as a three-level trial, with randomization at the school level [31]. Power is jointly determined by variability at the student, classroom, and school levels. Defining parameters include: intra-class correlation, number of sites (schools in this case), number of clusters per site (classrooms per school), cluster size (number of students per class). Including a level-3 covariate (cluster-level mean score at pretest) minimizes the conditional level-3 variance, increasing precision and power. Reasonable estimates for intraclass variability (5% between schools and 5% between classrooms; see [32]), and the proportion of explained variance due to the level-3 covariate (.60; see [33]) were used to estimate power requirements. Assuming 15 students per class (total of 2700 students), the minimum detectable effect at $p<.05$ is about .40 (Optimal Design for Longitudinal and Multilevel Research package; [34]). This is well below the results of pilot data, where the observed effects were closer to .60.

Analytic Strategy—For outcome measures, treatment main effects will be estimated for 1) physical activity; 2) math achievement on the WJ-III ACH; and 3) spelling achievement on the WJ-III ACH. Treatment groups will be compared across all measures, with the expectation that a given condition will outperform the other on content-related outcomes; in other words, the spelling group will outperform the math condition on spelling-related outcomes and the math condition will outperform students in the spelling group on math outcome measures. Controlling activity across these comparisons isolates the effect of content, and provides an estimate of its unique contribution to change in the academic-related outcomes. The no intervention control will serve as a no activity/no content comparison and be used to evaluate treatment effects on physical activity outcomes.

Schools were randomly assigned to treatment, threatening assumptions of independence and leading, potentially, to misestimated standard errors, heterogeneity of regression, and inflated Type I error rates [31, 35]. To minimize these threats and to account for the data's nested structure, multilevel analytic models will be used [31, 36, 37]. Unconditional multilevel models, estimated with a school-level covariate, (score on the relevant pretest), will be compared to conditional models with treatment indicated by a dummy-coded variable. Overall model fit will be estimated (standard fit criteria will be used – e.g., >.98 CFI, <.05 RMSEA), and statistically significant main treatment effects will be followed by pair-wise comparisons of condition-specific models (e.g., spelling group v. math group) and χ^2 difference tests to identify the sources of overall variation and to evaluate statistical significance [38].

DISCUSSION

This proposal was designed to address gaps in our understanding of physically active, academic lessons in the elementary school setting. The public health issue centers on the change in physical activity, with most public health researchers content with lessons that "do no harm." Unfortunately, the ability to disseminate these lessons will depend upon evidence from RCT that these lessons will enhance academic performance. Although the effectiveness of long-term interventions (2 year) on both physical activity and academic performance have been demonstrated, these measures do not mirror the weekly evaluations of performance that are the norm for most classrooms. Moreover, the existing research conflates physical activity with content of the lessons. This proposal addresses these limitations by assessing both proximal (weekly) and distal (6 month) academic and physical activity outcomes, with a test of key moderating factors. If the pilot data is replicated in this larger RCT, and significant effects are found for proximal outcomes and physical activity across all demographic categories, then these data will provide a strong inducement for dissemination and teacher implementation. As this exists outside of normal periods of physical activity (PE, recess, after school) successful dissemination of this approach during children's non-discretionary time will do much to achieve the public health goal of 60 min/day of activity for children. Finally, the design is unique in its ability to separate the impact of physical activity, in general, from the combination of physical activity and specific academic content. This will do much to guide the design of future, school-based interventions.

Acknowledgments

Funding. The project described was supported by Award Number 1R01HD070741 from the Eunice Kennedy Shriver National Institute of Child Health & Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health & Human Development or the National Institutes of Health.

Not Applicable

List of Abbreviations

BMI	body mass index
I-CAN	Initiatives for Children's Activity and Nutrition
MET	metabolic equivalent
MTS	momentary time sampling
MVPA	moderate-to-vigorous physical activity
PBC	perceived behavioral control
PE	physical education
RCT	randomized control trial
SES	socioeconomic status
TOT	time-on-task
WJ-III ACH	Woodcock Johnson Normative Update Tests of Achievement

References

1. Menschik D, Ahmed S, Alexander MH, Blum Rm. ADolescent physical activities as predictors of young adult weight. Arch Pediatr Adolesc Med. 2008; 162:29–33. [PubMed: 18180409]
2. Trudeau F, Shephard RJ. Physical education, school physical activity, school sports and academic performance. Int J Behav Nutr Phys Act. 2008; 5:10. [PubMed: 18298849]
3. Sy K, Nw G, Am K, Sl F, Dj A, Sl S, et al. Longitudinal changes in physical activity in a biracial cohort during adolescence. Med Sci Sports Exerc. 2000; 32:1445–54. [PubMed: 10949011]
4. Trost, S., Pate, RR., Sallis, JF., Freedson, PS., Taylor, WC., Dowda, M., et al. Age and gender differences in objectively measured physical activity in youth. Med Sci SPORTS Exerc. 2002. http://scholarworks.umass.edu/kinesiology_faculty_pubs/254

5. Carson RL, Castelli DM, Beighle A, Erwin H. School-Based Physical Activity Promotion: A Conceptual Framework for Research and Practice. Child Obes. 2014; 10:100–6. [PubMed: 24655311]
6. Donnelly JE, Greene JL, Gibson CA, Smith BK, Washburn RA, Sullivan DK, et al. Physical Activity Across the Curriculum (PAAC): A randomized controlled trial to promote physical activity and diminish overweight and obesity in elementary school children. Prev Med. 2009; 49:336–41. [PubMed: 19665037]
7. Gibson CA, Smith BK, Dubose KD, Greene JL, Bailey BW, Williams SL, et al. Physical activity across the curriculum: year one process evaluation results. Int J Behav Nutr Phys Act. 2008; 5:36. [PubMed: 18606013]
8. Stewart JA, Dennison DA, Kohl HW, Doyle JA. Exercise Level and Energy Expenditure in the TAKE 10!® In-Class Physical Activity Program. J Sch Health. 2004; 74:397–400. [PubMed: 15724566]
9. Honas JJ, Washburn RA, Smith BK, Greene JL, Donnelly JE. Energy Expenditure of the Physical Activity across the Curriculum Intervention. Med Sci Sports Exerc. 2008; 40:1501–5. [PubMed: 18614939]
10. Mahar MT, Murphy SK, Rowe DA, Golden J, Shields AT, Raedeke TD. Effects of a classroom-based program on physical activity and on-task behavior. Med Sci Sports Exerc. 2006; 38:2086–94. [PubMed: 17146314]
11. Reed KE, Warburton DER, Macdonald HM, Naylor PJ, McKay HA. Action Schools! BC: A school-based physical activity intervention designed to decrease cardiovascular disease risk factors in children. Prev Med. 2008; 46:525–31. [PubMed: 18377970]
12. Ahamed Y, Macdonald H, Reed K, Naylor P-J, Liu-Ambrose T, Mckay H. School-Based Physical Activity Does Not Compromise Children's Academic Performance. Med Sci Sports Exerc. 2007; 39:371–6. [PubMed: 17277603]
13. Mullender-Wijnsma MJ, Hartman E, de Greeff JW, Bosker RJ, Doolaard S, Visscher C. Improving Academic Performance of School-Age Children by Physical Activity in the Classroom: 1-Year Program Evaluation. J Sch Health. 2015; 85:365–71. [PubMed: 25877433]
14. Resaland GK, Aadland E, Moe VF, Aadland KN, Skrede T, Stavnsbo M, et al. Effects of physical activity on schoolchildren's academic performance: The Active Smarter Kids (ASK) cluster-randomized controlled trial. Prev Med. 2016; 91:322–8. [PubMed: 27612574]
15. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process. 1991; 50:179–211.
16. Bandura A. Social Cognitive Theory: An Agentic Perspective. Annu Rev Psychol. 2001; 52:1–26. [PubMed: 11148297]
17. Welk, GJ., Morrow, JRJ., Falls, HB. Fitnessgram Refence Guide. Dallas, TX: The Cooper Institute; 2002.
18. American College of Sports Medicine (ACSM). ACSM's guidelines for exercise testing and prescription. 7. Philadelphia: Lippincott Williams & Wilkins; 2006.
19. Montano, DE., Kasprzyk, D. The theory of reasoned action and the theory of planned behavior. In: Glanz, K.Rimer, BK., Lewis, FM., editors. Theory, research, and practice in health behavior and health education. San Francisco, CA: Josey Bass; 2002. p. 67
20. Welk GJ, Schaben JA, Morrow JR Jr. Reliability of accelerometry-based activity monitors: a generalizability study. Med Sci Sports Exerc. 2004; 36:1637–45. [PubMed: 15354049]
21. King GA, Torres N, Potter C, Brooks TJ, Coleman KJ. Comparison of activity monitors to estimate energy cost of treadmill exercise. Med Sci Sports Exerc. 2004; 36:1244–51. [PubMed: 15235333]
22. Treuth MS, Sherwood NE, Butte NF, Mcclanahan B, Obarzanek E, Zhou A, et al. Validity and Reliability of Activity Measures in African-American Girls for GEMS. Med Sci Sports Exerc. 2003; 35:532–9. [PubMed: 12618587]
23. McClain JJ, Abraham TL, Brusseau TA Jr, Tudor-Locke C. Epoch length and accelerometer outputs in children: comparison to direct observation. Med Sci Sports Exerc. 2008; 40:2080–7. [PubMed: 18981941]
24. Trost SG, Pate RR, Freedson PS, Sallis JF, Taylor WC. Using objective physical activity measures with youth: How many days of monitoring are needed? Med Sci Sports Exerc. 2000; 32:426–31. [PubMed: 10694127]

25. Vincent SD, Pangrazi RP. Does Reactivity Exist in Children When Measuring Activity Levels with Pedometers? Pediatr Exerc Sci. 2002; 14:56–63.
26. Gardenier NC, MacDonald R, Green G. Comparison of direct observational methods for measuring stereotypic behavior in children with autism spectrum disorders. Res Dev Disabil. 2004; 25:99–118. [PubMed: 15026089]
27. Grieco LA, Jowers EM, Bartholomew JB. Physically Active Academic Lessons and Time on Task: The Moderating Effect of Body Mass Index. Med Sci Sports Exerc. 2009; 41:1921–6. [PubMed: 19727020]
28. MacGinitie, W., MacGinitie, R., Maria, K., Dreyer, LG. Gates-MacGinitie Reading Tests. 4. Itasca, IL: Riverside Publishing Company; 2000.
29. MacGinitie WH, MacGinitie RK, Maria K, Dreyer LG, Hughes KE. Gates-MacGinities Reading Tests. Technical manual, Forms S & T. 2002
30. Woodcock, RW., McGraw, KS., Mather, N. Woodcock-Johnson III Tests of Achievement. Itasca, IL: Riverside; 2001.
31. Raudenbush, SW., Bryk, AS. Hierarchical Linear Models: Applications and Data Analysis Methods. SAGE; 2002.
32. Bloom HS, Richburg-Hayes L, Black AR. Using covariates to improve precision for studies that randomize schools to evaluate educational interventions. Educ Eval and Policy Anal. 2007; 29:30–59.
33. Konstantopoulos S. The Power of the Test for Treatment Effects in Three-Level Block Randomized Designs. J Res Educ Eff. 2008; 1:265–88.
34. Liu, X., Spybrook, J., Congdon, R., Martinez, A., Raudenbush, S. Optimal design for multi-level and longitudinal research. Ann Arbor, MI: University of Michigan, Survey Research Center of the Institute of Social Research; 2006.
35. Murray, DM. Design and Analysis of Group-randomized Trials. Oxford University Press; 1998.
36. Burchinal MR, Bailey DB, Snyder P. Using Growth Curve Analysis to Evaluate Child Change in Longitudinal Investigations. J Early Interv. 1994; 18:403–23.
37. Singer, JD., Willett, JB. Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence. Oxford University Press; USA: 2003.
38. Bovaird, JA. Multilevel structural equation models for contextual factors. In: Little, TD.Bovaird, JA., Card, NA., editors. Modeling contextual effects in longitudinal studies. Mahwah, NJ: Erlbaum; 2007. p. 149-182.

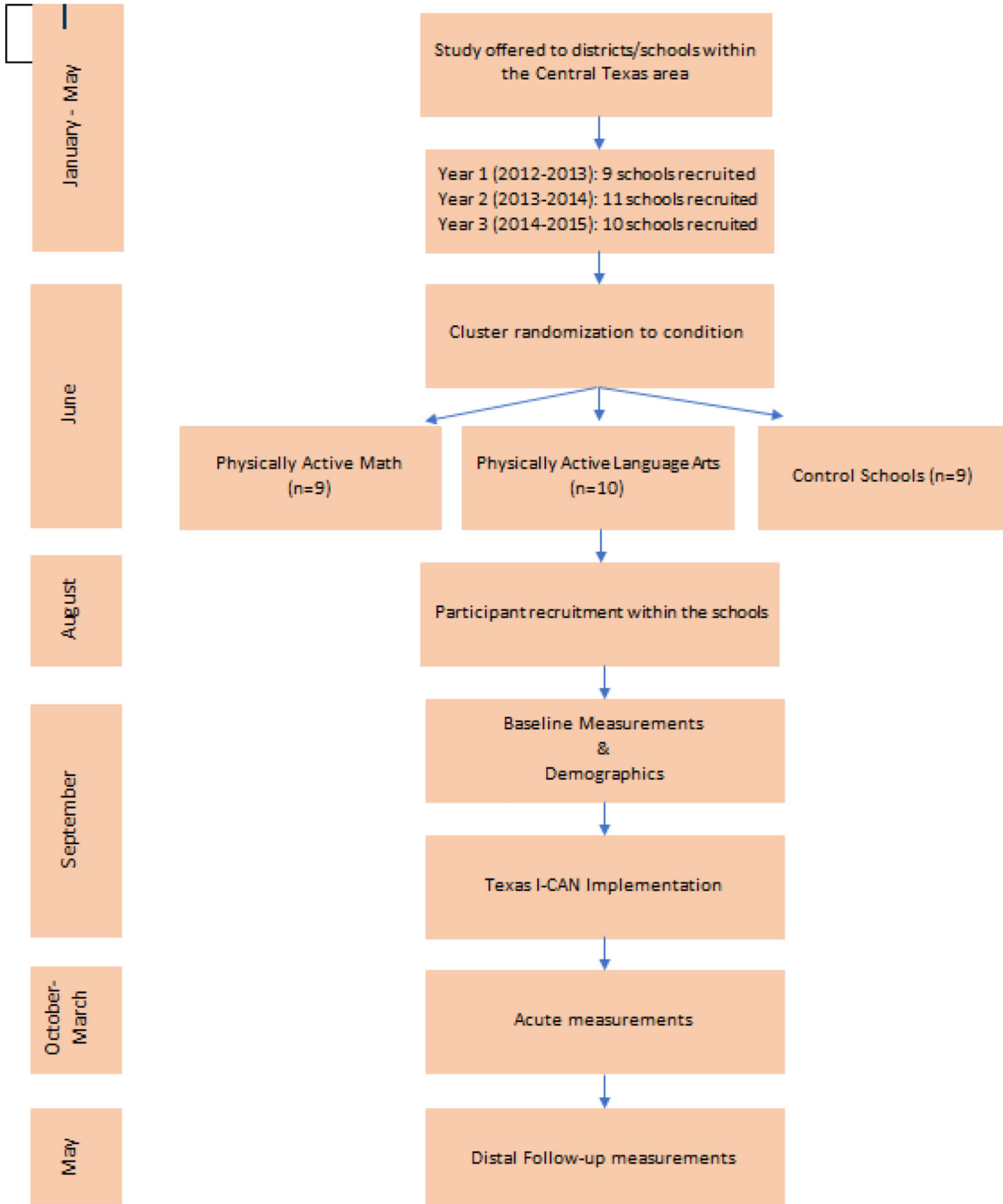

Fig. 1.
Flow diagram of study