What Works Clearinghouse™

ies INSTITUTE OF EDUCATION SCIENCES

February 2013

WWC Review of the Report "Evaluation of a Two-Year Middle-School Physical Education Intervention: M-SPAN"[1]

The findings from this review do not reflect the full body of research evidence on *Middle School Physical Activity and Nutrition (M-SPAN)*.

What is this study about?

The study examined the effect of the *Middle School Physical Activity and Nutrition (M-SPAN)* intervention on the physical activity level of middle school students.

For this 2-year study, 24 middle schools from six districts in southern California were stratified by school district and then randomly assigned to either *M-SPAN* or a comparison condition.

To assess students' physical activity levels and the content (referred to as "lesson context" in the article) of physical education (PE) classes, researchers observed students in PE classes on 11 randomly selected days for each school throughout the 2-year study period. Researchers documented the lesson content of the classes and observed a total of 1,849 lessons taught by 214 teachers (between seven and 14 teachers per school, with an average class size of 37.5 students).

The study assessed the effectiveness of *M-SPAN* by examining moderate-to-vigorous physical activity (MVPA), the amount of time students spent either walking or being very active, and other types of activities and PE lesson content across schools that received the *M-SPAN* training.[2]

Features of *Middle School Physical Activity and Nutrition (M-SPAN)*

Designed for middle school students, *M-SPAN* aims to increase physical activity in PE classes and reduce students' fat intake by encouraging healthy eating habits. During the 2-year study, *M-SPAN* trainers provided five 3-hour sessions of in-service training for intervention school teachers who volunteered to receive the professional development. The goal of the training, which included a package of curricular materials as well as goal-setting and modeling, was to increase students' moderate-to-vigorous physical activity (MVPA). Teachers received instruction and on-site coaching on setting goals for modifying PE at their schools; designing curricula that required active, health-related PE; and improving class management and instructional skills to enhance physical activity in class.[3]

What did the study find?

The study found that the *M-SPAN* intervention caused a statistically significant improvement in the amount of time students spent in MVPA, and the WWC confirms this study-level finding. The WWC calculated the *M-SPAN* intervention as improving the MVPA in schools by an average of 3 minutes per lesson (approximately 0.79 school standard deviation units) across the 2-year period of the study.

WWC Rating

The research described in this report meets WWC evidence standards without reservations

Strengths: This study is a well-implemented randomized controlled trial.

Cautions: The changes in observed MVPA (and other outcomes) may be in part due to (a) changes in MVPA in intervention schools, (b) high-activity students moving into the intervention schools or low-activity students moving out of the comparison schools, or (c) a combination of both effects. This analysis cannot separate these effects—it can only report on their combined impact.

Additionally, because the study analyzed school level data, the magnitude of the effects reported cannot be directly compared to the magnitude of an effect of an intervention that uses student-level data for the analysis.

Appendix A: Study details

McKenzie, T. L., Sallis, J. F., Prochaska, J. J., Conway, T. L., Marshall, S. J., & Rosengard, P. (2004). Evaluation of a two-year middle-school physical education intervention: M-SPAN. *Medicine & Science in Sports & Exercise, 36*(8),1382–1388.

Setting

The study was conducted in 24 public middle schools (grades 6–8) from six districts in Southern California.

Study sample

Participating schools had an average enrollment of 1,109 students. Among the student populations, 45% were non-White, and 39% were receiving free or reduced price meals.

Intervention group

In the schools assigned to the intervention, physical education (PE) teachers were offered the *Middle School Physical Activity and Nutrition (M-SPAN)* professional development, which included guidance on ways to improve their PE classes. During the 2-year study, *M-SPAN* trainers conducted five 3-hour sessions of in-service training for teachers in the intervention schools who volunteered to receive the professional development. The goal of the training, which included a package of curricular materials as well as goal-setting and modeling, was to increase students' moderate-to-vigorous physical activity (MVPA).

In a group setting, trainers provided teachers with sample curricular materials and helped them revise existing programs and instructional strategies. The sessions used didactic instruction and modeling/rehearsals as the main strategies. During the initial session, teachers set goals for modifying the PE lessons at their schools. These goals were revisited during the later sessions. Teachers were offered the chance to share with their peers the successful strategies that they had implemented at their schools. As noted below (see "Support for implementation"), teachers received on-site coaching to support program implementation.

Comparison group

The teachers in the comparison schools did not have access to the professional development sessions offered to the teachers in the intervention schools.

Outcomes and measurement

To assess students' physical activity levels and the content of PE classes, researchers observed students in PE classes on 11 randomly selected days for each school throughout the 2-year study period. In addition, researchers documented the content of the classes. Lesson content and student activity were assessed using SOFIT (System for Observing Fitness Instruction Time). The lesson content domain of SOFIT captures how class time is being spent at the time of observation. The student activity domain of SOFIT includes measures of how often students were observed to be engaged in a number of activity levels (such as sitting, walking, and very active). Researchers coded student activity levels for the SOFIT by randomly selecting four students in each classroom observation session and recording their activity every 20 seconds throughout the class time. For a more detailed description of these outcome measures, see Appendix B.

Support for implementation

M-SPAN trainers included three part-time, credentialed PE teachers, each with more than a decade of experience in public schools. These staff were trained by the study investigators to provide professional development to other PE teachers. To supplement the group education sessions, the trainers visited each school site twice per month in the first year and once per month in the second year to offer motivation, technical assistance, and feedback.

Reason for review

This study was identified for review by the WWC because it was suggested as a promising intervention through the WWC website's help desk.

Appendix B: Outcome measures for each domain

Student activity	
Lying down	Minutes per lesson spent lying down, measured using the System for Observing Fitness Instruction Times (SOFIT). These data were obtained by observing four randomly selected students in each classroom observation session and coding the frequency of the activity every 20 seconds.
Sitting	Minutes per lesson spent sitting, measured using SOFIT. These data were obtained by observing four randomly selected students in each classroom observation session and coding the frequency of the activity every 20 seconds.
Standing	Minutes per lesson spent standing, measured using SOFIT. These data were obtained by observing four randomly selected students in each classroom observation session and coding the frequency of the activity every 20 seconds.
Walking	Minutes per lesson spent walking, measured using SOFIT. These data were obtained by observing four randomly selected students in each classroom observation session and coding the frequency of the activity every 20 seconds.
Very active	Minutes per lesson spent being very active, measured using SOFIT. These data were obtained by observing four randomly selected students in each classroom observation session and coding the frequency of the activity every 20 seconds.
Moderate-to-vigorous physical activity (MVPA)	Minutes per lesson spent in MVPA, measured using SOFIT. This variable was obtained by summing the number of minutes spent in activity coded as either *walking* or *very active*.
Lesson content	
Management	Minutes per lesson spent on management, measured using SOFIT.
General knowledge	Minutes per lesson spent on general knowledge, measured using SOFIT.
Fitness knowledge	Minutes per lesson spent on fitness knowledge, measured using SOFIT.
Fitness activity	Minutes per lesson spent on fitness activity, measured using SOFIT.
Skill drills	Minutes per lesson spent on skill drills, measured using SOFIT.
Game play	Minutes per lesson spent on game play, measured using SOFIT.
Free play	Minutes per lesson spent on free play, measured using SOFIT.

Table Notes: Three additional outcomes were examined in this study, but are not included in this report because, as process measures, they focus on the implementation of the intervention rather than its outcomes. These include student enjoyment of and attendance at PE classes, teacher evaluation of group staff development sessions, and a teacher debriefing questionnaire.

Appendix C: Study findings for each domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Student activity								
Moderate-to-vigorous physical activity (MVPA)	Grades 6–8, Year 1	24 schools	18.9 (3.3)	17.0 (2.1)	1.90	0.66	+25	0.12
MVPA	Grades 6–8, Year 2	24 schools	19.4 (3.1)	16.9 (2.1)	2.50	0.91	+32	0.04
Domain average for student activity						**0.79**	**+28**	**Statistically significant**

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on school outcomes, representing the change (measured in standard deviations) in a school's outcome that can be expected if the school receives the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in a school's percentile rank that can be expected if the school is given the intervention. The WWC-computed average effect size is a simple average rounded to two decimal places; the average improvement index is calculated from the average effect size. The statistical significance of the study's domain average was determined by the WWC; the study is characterized as having a statistically significant positive effect because univariate statistical tests are reported for each outcome measure, the effect for at least one measure within the domain is positive and statistically significant, and no effects are negative and statistically significant, accounting for multiple comparisons.

Study Notes: A correction for multiple comparisons was needed and results in significance levels that differ from those in the original study. The *p*-values presented here were calculated by the WWC. The study author described a statistically significant impact on MVPA when pooling the information across all three time periods (baseline, Year 1, and Year 2), but not for the contrasts at each time period, so this report does not report the author's *p*-values in Appendix C. The WWC calculated the intervention group mean by adding the difference-in-differences adjusted estimate of the average impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttest means. Please see the *WWC Procedures and Standards Handbook* version 2.1 for more information.

Appendix D: Supplemental findings by domain

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Student activity								
Lying down	Grades 6–8, Year 1	24 schools	0.5 (0.1)	0.1 (0.1)	0.40	0.75	+27	0.08
Sitting	Grades 6–8, Year 1	24 schools	3.9 (1.8)	5.8 (2.8)	–1.90	–0.75	–27	0.08
Standing	Grades 6–8, Year 1	24 schools	12.2 (2.3)	12.1 (2.3)	0.10	0.00	0	1.00
Walking	Grades 6–8, Year 1	24 schools	13.5 (2.1)	12.4 (2.0)	1.10	0.52	+20	0.22
Very active	Grades 6–8, Year 1	24 schools	5.4 (1.5)	4.6 (1.2)	0.80	0.57	+22	0.18
Lying down	Grades 6–8, Year 2	24 schools	0.5 (0.1)	0.1 (0.1)	0.40	0.75	+27	0.08
Sitting	Grades 6–8, Year 2	24 schools	3.6 (1.8)	5.7 (2.4)	–2.10	–0.85	–30	0.05
Standing	Grades 6–8, Year 2	24 schools	13.7 (2.7)	12.4 (2.3)	1.30	0.48	+18	0.25
Walking	Grades 6–8, Year 2	24 schools	14.2 (2.3)	11.9 (2.3)	2.30	0.96	+33	0.03
Very active	Grades 6–8, Year 2	24 schools	5.3 (1.0)	5.0 (1.2)	0.30	0.25	+10	0.55
Lesson content								
Management	Grades 6–8, Year 1	24 schools	9.6 (1.0)	10.2 (2.0)	–0.60	–0.36	–14	0.39
General knowledge	Grades 6–8, Year 1	24 schools	2.0 (1.5)	2.1 (1.0)	–0.10	–0.08	–3	0.84
Fitness knowledge	Grades 6–8, Year 1	24 schools	0.1 (0.1)	0.1 (0.2)	0.00	0.00	0	1.00
Fitness activity	Grades 6–8, Year 1	24 schools	5.4 (3.8)	7.4 (2.2)	–2.00	–0.54	–20	0.20
Skill drills	Grades 6–8, Year 1	24 schools	3.9 (2.3)	2.6 (1.4)	1.30	0.80	+29	0.06
Game play	Grades 6–8, Year 1	24 schools	10.6 (2.6)	9.9 (3.6)	0.70	0.07	+3	0.87
Free play	Grades 6–8, Year 1	24 schools	4.8 (1.7)	2.9 (2.6)	1.90	0.65	+24	0.12
Management	Grades 6–8, Year 2	24 schools	10.9 (2.5)	10.9 (2.3)	0.00	–0.01	–1	0.97
General knowledge	Grades 6–8, Year 2	24 schools	1.9 (0.9)	1.7 (1.0)	0.20	0.17	+7	0.68

Domain and outcome measure	Study sample	Sample size	Mean (standard deviation)		WWC calculations			*p*-value
			Intervention group	Comparison group	Mean difference	Effect size	Improvement index	
Fitness knowledge	Grades 6–8, Year 2	24 schools	0.0 (0.0)	0.2 (0.5)	–0.20	–0.55	–21	0.19
Fitness activity	Grades 6–8, Year 2	24 schools	4.4 (3.4)	7.7 (2.9)	–3.30	–0.94	–33	0.03
Skill drills	Grades 6–8, Year 2	24 schools	3.5 (3.0)	1.8 (1.0)	1.70	0.92	+32	0.03
Game play	Grades 6–8, Year 2	24 schools	11.5 (5.2)	8.9 (5.5)	2.60	0.49	+19	0.25
Free play	Grades 6–8, Year 2	24 schools	5.9 (2.9)	3.9 (6.0)	2.00	0.79	+29	0.07

Table Notes: For mean difference, effect size, and improvement index values reported in the table, a positive number favors the intervention group and a negative number favors the comparison group. The effect size is a standardized measure of the effect of an intervention on school outcomes, representing the change (measured in standard deviations) in a school's outcome that can be expected if the school receives the intervention. The improvement index is an alternate presentation of the effect size, reflecting the change in a school's percentile rank that can be expected if the school is given the intervention.

Study Notes: A correction for multiple comparisons was needed and results in significance levels that differ from those in the original study. The *p*-values presented here were calculated by the WWC, and none of the contrasts were found to be statistically significant after adjusting for multiple comparisons. The author did not report inferential tests of any of the contrasts presented in Appendix D. The WWC calculated the intervention group mean by adding the difference-in-differences adjusted estimate of the average impact of the program (i.e., difference in mean gains between the intervention and comparison groups) to the unadjusted comparison group posttests means. Please see the *WWC Procedures and Standards Handbook* version 2.1 for more information.

Endnotes

[1] Single study reviews examine evidence published in a study (supplemented, if necessary, by information obtained directly from the author[s]) to assess whether the study design meets WWC evidence standards. The review reports the WWC's assessment of whether the study meets WWC evidence standards and summarizes the study findings following WWC conventions for reporting evidence on effectiveness. This study was reviewed using the single study review protocol, version 2.0. The WWC rating applies only to the results that were eligible under this topic area and met WWC standards without reservations or met WWC standards with reservations, and not necessarily to all results presented in the study.

[2] Three additional outcomes were examined in this study but are not included in this report because, as process measures, they focus on the implementation of the intervention rather than its outcomes. These include student enjoyment of and attendance at PE classes, teacher evaluation of group staff development sessions, and a teacher debriefing questionnaire. In addition, the author presented sub-group estimates for the MVPA outcome by gender; however, there was insufficient information to include those estimates in this report.

[3] The *M-SPAN* intervention also included an environmental, policy, and social marketing intervention to encourage healthy eating (reducing fat intake). The authors note that the because of these additional components, the PE component of this intervention may have taken place in a favorable implementation context.

Recommended Citation

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, February). *WWC review of the report: Evaluation of the two-year middle-school physical education intervention: M-SPAN*. Retrieved from http://whatworks.ed.gov.

Glossary of Terms

Attrition Attrition occurs when an outcome variable is not available for all participants initially assigned to the intervention and comparison groups. The WWC considers the total attrition rate and the difference in attrition rates across groups within a study.

Clustering adjustment If intervention assignment is made at a cluster level and the analysis is conducted at the student level, the WWC will adjust the statistical significance to account for this mismatch, if necessary.

Confounding factor A confounding factor is a component of a study that is completely aligned with one of the study conditions, making it impossible to separate how much of the observed effect was due to the intervention and how much was due to the factor.

Design The design of a study is the method by which intervention and comparison groups were assigned.

Domain A domain is a group of closely related outcomes.

Effect size The effect size is a measure of the magnitude of an effect. The WWC uses a standardized measure to facilitate comparisons across studies and outcomes.

Eligibility A study is eligible for review if it falls within the scope of the review protocol and uses either an experimental or matched comparison group design.

Equivalence A demonstration that the analysis sample groups are similar on observed characteristics defined in the review area protocol.

Improvement index Along a percentile distribution of students, the improvement index represents the gain or loss of the average student due to the intervention. As the average student starts at the 50th percentile, the measure ranges from –50 to +50.

Multiple comparison adjustment When a study includes multiple outcomes or comparison groups, the WWC will adjust the statistical significance to account for the multiple comparisons, if necessary.

Quasi-experimental design (QED) A quasi-experimental design (QED) is a research design in which subjects are assigned to intervention and comparison groups through a process that is not random.

Randomized controlled trial (RCT) A randomized controlled trial (RCT) is an experiment in which investigators randomly assign eligible participants into intervention and comparison groups.

Single-case design (SCD) A research approach in which an outcome variable is measured repeatedly within and across different conditions that are defined by the presence or absence of an intervention.

Standard deviation The standard deviation of a measure shows how much variation exists across observations in the sample. A low standard deviation indicates that the observations in the sample tend to be very close to the mean; a high standard deviation indicates that the observations in the sample are spread out over a large range of values.

Statistical significance Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups. The WWC labels a finding statistically significant if the likelihood that the difference is due to chance is less than 5% ($p < 0.05$).

Substantively important A substantively important finding is one that has an effect size of 0.25 or greater, regardless of statistical significance.

Please see the WWC Procedures and Standards Handbook (version 2.1) for additional details.