



**Globaloria Replication Study: An Examination
of the Relationships between Globaloria
Participation and Student Achievement in
Year 4 of the West Virginia Pilot
Implementation**

Hsiang-Yeh Ho, Ph.D.
Jessica Gore, M.A.
Kristine Chadwick, Ph.D.

Tennessee Office

One Vantage Way, Suite C-200
Nashville, TN 37228
615.565.0101

West Virginia Office

500 Virginia Street E, Suite 500
Charleston, WV 25301
800.624.9120

info@edvantia.org

Founded in 1966, Edvantia is a not-for-profit corporation with primary offices in Tennessee and West Virginia. Edvantia works in partnership with clients to provide workable solutions to the issues facing education today. Our comprehensive services are grounded in research and best practices and delivered by a team of social scientists, and former teachers, administrators, and state education agency leaders who are proven experts in program evaluation and school improvement.

For information about Edvantia research, products, or services, contact



500 Virginia Street East, Suite 500, Charleston, WV 25301 • 304.347.0400 • fax 304.347.0487
One Vantage Way, Suite C-200, Nashville, TN 37228 • 615.565.0101 • fax 615.565.0112
info@edvantia.org • www.edvantia.org

© 2012 by Edvantia, Inc.

This publication was developed under a contract with the World Wide Workshop Foundation. However, its contents do not necessarily represent the policy of the Foundation, and you should not assume endorsement by the Foundation.

Edvantia is an equal employment opportunity/affirmative action employer.

Table of Contents

Executive Summary	v
Introduction	1
The World Wide Workshop Foundation	1
Globaloria: Underlying Theory and Structure	1
Globaloria-WV	3
Brief Research Review for Globaloria-WV	3
Purpose of this Study	4
Methods	6
Sample Selection	6
Analytic Plan	9
Results	11
Mathematics	11
Science	12
Reading	15
Social Studies	16
Key Cross-Year Findings	18
Conclusions and Recommendations	19
Cross-Year Comparison	19
Recommendations	20
Limitations	21
References	22

Appendices

- A: School-Level Globaloria Implementation and Learning Outcomes Logic Model
- B: Participating and Matched Comparison School Selection Criteria
- C: Propensity Score Matching Steps and Procedures
- D: Hierarchical Linear Modeling: Impact Model

List of Tables

Table 1. Means and Standard Deviations of School-Level Covariates by Participation	6
Table 2. Number of Globaloria Participating Students by School Before and After Matching	7
Table 3. Means, Standard Deviations, and Balance Diagnosis after Matching	8
Table 4. Results of Two-Level HLM: Mathematics	11
Table 5. Results of Two-Level HLM: Science	12
Table 6. Results of Two-Level HLM: Reading.....	15
Table 7. Results of Two-Level HLM: Social Studies	17

List of Figures

Figure 1. The Globaloria platform and its curricular units guide students and educators through a multidimensional blended learning process.	2
Figure 2. Relationship between participation and math proficiency on science subtest.	14
Figure 3. Relationship between participation and reading proficiency on science subtest.	14
Figure 4. Relationship between participation and math proficiency on reading subtest.	16

Executive Summary

This study examined the relationships between Globaloria participation and student achievement, measured by four subtests of West Virginia's standardized assessment test, the WESTEST2: mathematics, science, reading, and social studies. The study expands upon the previous two studies conducted by Chadwick and Gore (2010; 2011) in two major ways: (1) increases the sample size and (2) uses a modified, more rigorous quasi-experimental methodology that allows researchers to take into consideration within- and between-school variations in student achievement.

Propensity score matching (PSM) methods were used to create a counterfactual group that was similar to the group of Globaloria students in various aspects of student-level characteristics (e.g., grade level, gender, income status) and school-level characteristics (e.g., percent of students achieving reading proficiency on the 2010 WESTEST2, percent of students achieving math proficiency on the 2010 WESTEST2, school enrollment, and percent of low-income students). Due to the complex sampling design (i.e., students nested within schools), two-level hierarchical linear modeling (HLM) was used to analyze the data where students (level 1) were nested within schools (level 2). This analytic design was more robust than the design used in the previous Globaloria pilot and replication studies (Chadwick & Gore, 2010, 2011). HLM provided the advantages of taking within- and between-school variations in student achievement into account. Given the limited sample sizes of the racial subgroups, only White students were included in the analyses. Following is a summary of the key HLM findings for each subject area assessed by the WESTEST2:

- Students who participated in Globaloria scored significantly higher than nonparticipating students on the **mathematics** subtest.
- While there were no significant differences between participating and nonparticipating students' **science** scores overall, the influence of Globaloria participation on science scores seemed to be affected by the percentage of students achieving proficiency in mathematics at the school. Specifically, participating students in schools with low mathematics proficiency scored significantly higher on the science subtest than their nonparticipating counterparts.
- Students who participated in Globaloria scored significantly higher than nonparticipating students on the **reading** subtest. Similar to the finding for science, the effects of Globaloria participation were stronger in schools with a low percentage of students achieving mathematics proficiency.
- No significant differences were found between the scores of participants and nonparticipants on the **social studies** subtest.

Taken together, the findings suggest that Globaloria participation positively affects students' math and reading achievement, and to a lesser extent, science achievement, as measured by the WESTEST2. Additionally, it seems that school-level contextual factors may be important in determining the extent to which participation will affect science and reading achievement.

While it is difficult to compare the findings across the three studies (pilot study, replication study, and current study using a HLM design), researchers compared the key findings to the extent possible. In the pilot study, Globaloria participants outperformed their comparison counterparts in science and social studies achievement; however, only the difference for science was upheld in the replication study. In the current study, Globaloria participants outperformed their comparison counterparts in mathematics and reading, which is different from the findings of the previous two studies. While there continued to be some influence of Globaloria participation on science achievement, in this year's analyses the influence was observed only in schools with low overall

mathematics achievement. It is likely that the change in findings is due to the increased sample size, improved matching procedures, and the robust HLM design that takes within- and between-school variations in student achievement into consideration.

Based on these findings, the following recommendations are provided for future research on Globaloria:

- The World Wide Workshop Foundation (the Workshop) is encouraged to re-examine the logic model suggested by Edvantia in 2008 (Knestis, 2008) to identify “mediators” to help explain the processes that link program participation with student outcomes. Understanding these mediators is key to successful program scale-up.
- The Workshop WWWF program staff are encouraged to consider replicating the current study methodology using a more diverse sample of students than in the current study. This likely would require drawing upon Globaloria implementation sites other than West Virginia, a rather demographically homogenous state.
- Researchers interested in studying Globaloria are encouraged to consider a longitudinal study with middle school students to investigate the extent to which Globaloria participation may have long-term effects on students, both in terms of achievement outcomes and college- and career-readiness outcomes.
- It is recommended that future studies bring together research that has been conducted on teacher characteristics and behaviors (i.e., proximal moderators), and the effects of Globaloria participation on student achievement outcomes.
- Findings from this study suggest that “school context” matters. It is recommended that future research include examination of factors that are malleable (e.g., educator practice and interaction with students) so these factors can be incorporated into the Globaloria model as it evolves.
- It is important that future studies investigate “self-selection” factors that may be playing a part in the evidenced relationships between participation and achievement. Most students who participate in Globaloria are likely to be on track to graduate; if they were not, they would not have room in their course schedule for an elective on game design. They may have stronger interests in mathematics, technology, and science than students who do not participate in Globaloria, and the participating students may be more motivated to learn new things than their nonparticipating counterparts.
- Future studies should examine the relationship between the type of game (e.g., game topic and/or content) and students’ performance on the related area of the WESTEST2.

Introduction

The World Wide Workshop Foundation (herein referred to as the Workshop) contracted with Edvantia to conduct a replication evaluation of the *Globaloria Pilot Study: The Relationship of Globaloria Participation and Student Achievement* (Chadwick & Gore, 2010) and the *Globaloria Replication Study: Examining the Robustness of Relationships between Globaloria Participation and Student Achievement* (Chadwick & Gore, 2011). The study documented in this report made two improvements to the first two studies: (1) increases the sample size, and (2) uses a modified, more rigorous quasi-experimental methodology that allows researchers to take into consideration within- and between-school variations in student achievement. This report used the Year 4 data from the Globaloria West Virginia Pilot Implementation (heretofore referred to as *Globaloria-WV*). The main purpose of this study seeks to provide empirical evidence to support the effectiveness of *Globaloria-WV* on student achievement, as measured by West Virginia's standardized achievement test, the WESTEST2.¹

The World Wide Workshop Foundation

The Workshop, founded in 2004, is a non-profit organization that invents social media and digital technology application to help youth and educators participate as leaders in the global knowledge economy. The Workshop partners with “forward-thinking leaders, corporations, school systems, universities, foundations, and research centers worldwide to enrich existing formal and non-formal education with the latest technology and innovative learning opportunities.” The mission of the Workshop is to “develop applications for learning with technology that combines game mechanics and social networking to empower youth to be inventors and leaders in the global knowledge economy.” The Workshop's programs are designed to “transform education by connecting youth to learning, community engagement, and economic development through game production” (World Wide Workshop Foundation, 2012).

Globaloria: Underlying Theory and Structure

The Workshop created Globaloria, a social network for learning web-based game design and simulation production to address the two digital divides encountered by poor and underserved communities in the United States and worldwide. The first digital divide is defined by issues with access to high-speed Internet. The second divide, coined as “digital literacy,” is the ability to create, not just consume, digital media (Knestis, 2008). The Globaloria program is based on the constructivist educational philosophy (Harel, 1991; World Wide Workshop Foundation, 2008). According to Harel (1991), the constructionist approach to learning operates from the view that “building knowledge structures (‘in the head’) goes especially well when the subject is engaged in building material structures (‘in the world’).” Through this approach, “children learn how to learn, and they learn how to think about thinking” (Harel, 1991). This is accomplished through “publicly shared, long-term projects that are complex, computational, immersive, and innovative” in which students learn by doing (World Wide Workshop Foundation, 2008). Research has shown that constructivist programs result in deeper forms of learning, cognitive integration, and improved approaches to learning (Rogers, Pertosino, Huebner, & Hacsí, 2000).

¹ The WESTEST2 provided a valid measure of the effects of the Globaloria program because the WESTEST2 was designed to measure 21st century skills, as represented by the West Virginia content standards and objectives (CSOs), while Globaloria is designed to increase them.

The goal of the Globaloria program is to create technology-based educational opportunities through a series of virtual learning networks (World Wide Workshop Foundation, 2008, 2009). That is, Globaloria allows participants to interact with games, puzzles, and creative tools, while also thinking as game developers (World Wide Workshop Foundation, 2009, 2012). When participants create their own games, they are encouraged to share their work on the shared wiki platform (World Wide Workshop Foundation, 2008). Through construction, interaction, and play, the Globaloria program empowers youth to be productive, successful 21st century citizens, by fostering the following essential six contemporary learning abilities (CLAs) inherent in digital literacy (Harel, Oliver, & Sullivan, 2010):

- Invention, progression, completion of an original project: program an educational game, wiki, or simulation
- Project-based learning in Web 2.0² environments and processing complex project management (programmable wiki systems)
- Producing, programming, publishing, and distributing interactive purposeful digital media
- Social learning, participation, and exchange
- Information-based learning, search, and exploration
- Thoughtful surfing of websites and web applications.

Globaloria is the platform and curriculum that takes students through a rich learning process, as depicted in Figure 1. A complete review of this framework is available at the Workshop website (<http://www.worldwideworkshop.org/reports>). A brief review is also available in Chadwick and Gore (2011).



Figure 1. The Globaloria platform and its curricular units guide students and educators through a multidimensional blended learning process.

² A Web 2.0 environment is a shift in the way users think about technology from consumers to creators.

Globaloria is a “rigorous turn-key instructional solution for STEM learning,” comprising a year-long academic curriculum, game design and programming tutorials, game-content resources, and virtual support systems for educators and youth (<http://www.worldwideworkshop.org/>). Working independently and in small teams, students drive the design process, taking an original idea to final game product. No prior web design or programming skills are needed. Learning by doing, students are educated in both technical and computational skills and in content knowledge that prepares them for college-level studies, as well as for digital citizenship and careers in the global knowledge economy (World Wide Workshop Foundation, 2012). It is hypothesized that through participation in the Globaloria program, students develop 21st century skills in digital literacy and social media while gaining a deeper understanding of curricular areas, such as science, mathematics, health, and global issues (Edvantia, 2008), which is essential for success in the 21st century (Harel et al., 2010).

Nevertheless, through participating in Globaloria, students are also involved in and exposed to a variety of academic subjects and social issue topics, such as science, math, language arts, architecture, art, climate change, ecology, water, community services, technology skills, peace, and more (Harel et al., 2010). Therefore, the potential impacts of Globaloria participation are likely to go beyond the scope of STEM education and learning (Chadwick & Gore, 2011; Reynolds & Chiu, 2012).

Globaloria-WV

In August 2007, the Workshop deployed Globaloria into classrooms throughout West Virginia. It was the first statewide implementation of a social learning network using game design (Chadwick & Gore, 2010, 2011; World Wide Workshop Foundation, 2009). Since then, *Globaloria-WV* has included thousands of students from middle schools, high schools, community colleges, and alternative education programs across the state. Educators implement Globaloria as an elective, a high school completer course, and as an integrated program within the school core curriculum, aligning the program with Global21 state standards and objectives in English/language arts, mathematics, science, and 21st century skills (World Wide Workshop Foundation, 2012). As a first-of-its-kind statewide network for learning, this model demonstrates the potential for Globaloria to be implemented on a large scale, integrated at multiple grade levels, across the public school system and higher education.

To evaluate the cognitive, behavioral, and affective impact of the Globaloria program, the Workshop researchers have used a variety of evaluation methods, including the use of surveys; tracking of activities and behaviors; case studies; evaluations of work products; evaluations of wiki participation; interviews, conference calls, and e-mail exchanges with educators; in-person visits; and videos and transcriptions from site visits. Information from these data collection measures informed a comprehensive evaluation of the impact of Globaloria and the refinement of the program over time (World Wide Workshop Foundation, 2012). This section summarizes the key findings from previous evaluations and activities. For a complete list of prior studies and reports related to *Globaloria-WV*, please visit <http://www.worldwideworkshop.org/reports>.

Brief Research Review for Globaloria-WV

In December 2008, Edvantia researchers developed a research agenda based on the review of extant data (e.g., project descriptions, evaluation reports, participant artifacts) to guide future study of the Globaloria education intervention (Knestis, 2008). The agenda included a logic model describing the associations between program activities and different phases of program outcomes

(see Appendix A for the logic model). To date, multiple research and evaluation studies have been conducted to examine the various relationships between program components and program outcomes depicted in the logic model. Findings have been positive for the program. Specifically, during the second year of the pilot study, research indicated that middle school student attitudes toward Globaloria activities and self-reported knowledge increased, especially for the more constructionist, effortful, and difficult activities (Reynolds, Scialdone, & Harel, 2010). In the Year 3 evaluation, the goal was to expand the evaluation of *Globaloria-WV* to include not only the impact of Globaloria on contemporary learning abilities and teacher professional development (Whitehouse, 2009), but also its effect on student achievement, performance, and the relationship among classroom ecology, game design pedagogy, and Globaloria learning (Harel et al., 2010). These evaluation studies provided preliminary evidence that participation in Globaloria may have positive effects on student performance and academic achievement. Specifically, researchers found that Globaloria students scored moderately higher on five out of six assessment measures (three unit tests, a semester final, course average, and course grade) (Harder, Stuart, & Anthony, 2010; Nicholson, Alley, Green, & Lawson, 2009; World Wide Workshop Foundation, 2012).

In recent reports, Reynolds and colleagues extended their work to examine how location-level factors (e.g., educator experience with the program, educator training, motivation and expertise, and teamwork models) might be contributing to variations in student outcomes (Reynolds & Kim, 2012). This research is currently underway, and findings will provide a valuable framework for predicting classes that may require additional support and guidance.

Purpose of this Study

Edvantia has worked with the Workshop for the past four years to explore the effectiveness of *Globaloria-WV*. In 2010, Edvantia conducted a pilot study of the effects of Globaloria participation on student achievement. This study provided preliminary data concerning the performance of students participating in Globaloria relative to the Global21 standards, as measured by the WESTEST2, in four key content areas: math, reading/language arts, science, and social studies. Findings indicated that participants in Globaloria performed significantly better on the 2009 WESTEST2 science and social studies subtests than those who did not participate in the program (Chadwick & Gore, 2010); however, this study was limited by its small sample size.

In 2011, Edvantia conducted a replication study to gather further support for the pilot findings. In the replication study, researchers increased the sample size and used a more robust matching procedure (propensity score matching [PSM]) than was possible for the pilot study. Researchers found that students who participated in the Globaloria program scored at least slightly higher than comparison students on all four subsections of the 2010 WESTEST2; however, a statistical *significant* difference was found only in the WESTEST2 science domain after controlling for the pretest scores. This finding suggested that Globaloria participation was positively related to student science performance, but was not related—or only loosely related—to other subjects tested on the WESTEST2 (Chadwick & Gore, 2011).

Although the Globaloria program has shown promise for improving digital literacy in West Virginia students, more confidence in the relationships between program participation and achievement outcomes would result if similar outcomes were found with a larger sample size. The purpose of this study is to replicate the *Globaloria Pilot Study: The Relationship of Globaloria Participation and Student Achievement* (Chadwick & Gore, 2010) and the *Globaloria Replication Study: Examining the Robustness of Relationships between Globaloria Participation and Student Achievement* (Chadwick & Gore, 2011) using the Year 4 data from the *Globaloria-WV* pilot

implementation and a modified, more rigorous quasi-experimental methodology. Due to the complex sampling design (i.e., students nested within schools), two-level hierarchical linear modeling (HLM) was used to analyze the data where students (level 1) were nested within schools (level 2). This analytic approach was greatly improved from previous Globaloria studies (Chadwick & Gore, 2010, 2011). Previously, multivariate analysis of covariance (MANCOVA) was used to examine the associations between Globaloria participation and student achievement outcomes. MANCOVA was a favorable approach when the study sample size was small; yet, this analytic method did not allow the researchers to take into account within- and between-school variations in student achievement outcomes. In other words, the MANCOVA assumes that each student's achievement is independent from others; however, in the real world, students from the same schools are more similar than students from different schools. Given the adequate sample size available for the Year 4 analyses, the more robust approach (e.g., HLM) provided the advantages of taking these within- and between-school variations in student achievement into account.

This study seeks to expand the previous studies and provide additional data concerning the performance of students participating in *Globaloria-WV* relative to the Global21 standards, as measured by the 2011 WESTEST2. See Appendix A for how this study relates to the previously developed logic model and research agenda (Knestis, 2008).

Methods

In Year 4, 42 schools and education institutions participated in *Globaloria-WV*. Of those, Edvantia evaluators included all traditional middle and high schools (9 and 24, respectively) in the replication study. To create a counterfactual group, evaluators used propensity score matching (PSM) (Stuart & Rubin, 2007) to select a group of comparison students from similar schools who were similar to participating students on key variables prior to participation. In the field of education and evaluation where random assignment is not always feasible or ethical, PSM provides advantages to examine the effects of interventions such as Globaloria (Stuart, 2010).

Sample Selection

The PSM consisted of two stages. The first stage involved selecting a group of comparison schools³ that were similar to the Globaloria schools on the following four key school-level characteristics: (1) percent of students achieving reading proficiency on the 2010 WESTEST2, (2) percent of students achieving math proficiency on the 2010 WESTEST2, (3) school enrollment, and (4) percent of students qualifying for free or reduced-price meal status.⁴ Free or reduced-price meal status was used as a proxy for income (i.e., students who qualified for free or reduced-price meal status were considered to be low-income students). The matching process ensured that the selected comparison schools had similar demographic and achievement characteristics to participating schools before participation in Globaloria. See Appendix B for the list of participating and matched comparison schools. As shown in Table 1, results of independent samples *t*-tests showed that the selected comparison schools were not statistically different from the participating schools in these four school-level covariates.

Table 1. Means and Standard Deviations of School-Level Covariates by Participation

School-Level Covariates	Participating (<i>n</i> = 33)		Comparison (<i>n</i> = 33)		Independent Samples <i>t</i> -test		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	<i>p</i>
% Math Proficiency	40.87	10.79	41.30	112.00	0.15	64	0.88
% Reading Proficiency	45.91	9.22	45.98	10.56	0.03	64	0.98
% Low-income Students	49.31	11.33	48.48	16.63	-0.24	64	0.81
School Enrollment	671.24	349.48	707.36	478.31	0.35	64	0.73

The second stage was to select a group of comparison students from the identified schools to match with participating students on the following key student-level attributes: gender, race, low-income status, Limited English Proficiency (LEP), disability, grade level, and 2010 WESTEST2 scale scores. For this stage of matching, the Workshop provided Edvantia researchers with a list of

³ Evaluators obtained a complete list of public middle and high schools in West Virginia from the West Virginia Department of Education website (<http://wveis.k12.wv.us/nclb/pub/>). At the request of the Workshop, all traditional middle and high schools were included. Colleges and alternative schools were excluded.

⁴ These school-level as well as student-level characteristics were used because they were commonly used in PSM literature (Stuart, 2010).

818 participants for which project staff had obtained parental consent⁵ for participation in the study. Data for the covariates were provided by the West Virginia Department of Education. Since PSM requires full information to proceed with the matching, multiple imputations were used to manage missing values⁶ (Hill, 2004). Missing patterns were examined before the imputation.⁷ After screening and imputing missing values, propensity scores were obtained by using logistic regression (Austin, 2011). The final sample included 701 participating and 701 comparison students. Table 2 displays the total number of participating students before and after PSM by school. Details for PSM steps and procedures are reported in Appendix C.

Table 2. Number of Globaloria Participating Students by School Before and After Matching

School	Initial Participating Students	Final Sample
Braxton County High School	10	10
Buffalo High School	6	6
Cameron High School	4	4
Capital High School	25	25
Chapmanville Middle School	20	16
Chapmanville Regional High School	14	11
Doddridge County High School	18	18
Dunbar Middle School	3	1
Eastern Greenbrier Middle School	24	16
Fayetteville High School	26	24
George Washington High School	19	18
Greenbrier East High School	16	13
Greenbrier West High School	22	19
Hurricane High School	21	20
Kasson Middle School	17	11
Liberty High School	11	9
Logan Middle School	11	10
Logan Senior High School	10	10

⁵ All students who participate in the Globaloria program are required by the Workshop to return parental consent to participate in the program and all research activities related to the program; students who do not return consent cannot participate.

⁶ Attrition analysis was conducted to examine the differences between samples without full information ($n = 404$) and samples with full information ($n = 362$). Results showed that the samples without full information were more likely to be those who were male, minority, low income, and have lower reading scores before participation. Removing these cases is likely to bias the estimations of treatment effect; therefore it is more appropriate to retain all cases and use multiple imputation to manage missing values (Hill, 2004).

⁷ Of the 818 Globaloria students, 41 were removed due to incorrect West Virginia Education Information System identification numbers. Eight students were removed from the sample because they were missing more than two of the four 2010 WESTEST2 scale scores. Of the remaining 769 Globaloria students, 35 were removed due to missing data on demographic variables. Finally, 34 Globaloria students were removed from the sample because the PSM procedure did not produce good matches for these individuals; these students had higher 2010 WESTEST2 scores for all four subscales.

School	Initial Participating Students	Final Sample
Man High School	9	9
Man Middle School	11	11
Oak Glen High School	14	11
Oak Hill High School	17	14
Philip Barbour High School	53	38
Riverside High School	137	126
Sandy River Middle School	53	38
South Charleston Middle School	20	12
South Harrison High School	19	18
Spring Valley High School	113	98
Tygarts Valley High School	20	18
Webster County High School	24	24
Weir High School	11	10
Western Greenbrier Middle School	11	9
Wheeling Park High School	29	24
TOTAL	818	701

As displayed in Table 3, results of independent samples *t*-tests, chi-square tests, and balance diagnostic criteria, suggested by Rubin (2001), indicated that the selected comparison students were similar to Globaloria students in all aspects of student-level covariates. See Appendix C for the descriptions of balance diagnostic criteria.

Table 3. Means, Standard Deviations, and Balance Diagnosis after Matching (n = 701)

Variables	Participating		Comparison		Balance Diagnosis		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Standardized Mean Difference	<i>p</i> Value	Variance Ratio
Propensity Score	0.06	0.03	0.06	0.03	-0.01	1.00	1.00
Pretest Covariates							
2010 Math	654.74	45.22	654.66	49.47	0.00	0.98	0.91
2010 Reading	481.60	43.86	481.30	47.25	0.01	0.90	0.93
2010 Science	604.76	47.09	606.38	50.70	-0.03	0.54	0.93
2010 Social Studies	408.11	32.40	407.43	35.39	0.02	0.71	0.92
Demographic Covariates	Participating Percent		Comparison Percent		Chi-square test	<i>p</i> Value	% Ratio
Student Grade Level					$\chi^2 = 32.93 (6)$	<.001	
Sixth Graders	1.0%		1.0%				1.00
Seventh Graders	8.3%		7.0%				1.19
Eighth Graders	8.4%		8.7%				0.97

Demographic Covariates	Participating Percent	Comparison Percent	Chi-square test	p Value	% Ratio
Ninth Graders	7.1%	2.6%			2.73
Tenth Graders	21.1%	29.7%			0.71
Eleventh Graders	24.0%	27.4%			0.88
Twelfth Graders	30.1%	23.7%			1.27
Male	63.8%	67.3%	$\chi^2 = 1.97 (1)$	0.16	0.95
Low-Income	42.7%	45.5%	$\chi^2 = 1.16 (1)$	0.28	0.94
LEP	0.1%	0.0%	$\chi^2 = 1.00 (1)$	0.32	1.00
Disability	5.7%	4.7%	$\chi^2 = 0.71 (1)$	0.40	1.21
High School Students	82.3%	83.3%	$\chi^2 = 0.25 (1)$	0.62	0.99
Black Students	3.7%	2.4%	$\chi^2 = 1.94 (1)$	0.16	1.54
Asian Students	0.7%	0.6%	$\chi^2 = 0.11 (1)$	0.74	1.17

Analytic Plan

Of specific interest in this study was the relationship between students' participation status (whether or not they participated in the Globaloria program) and WESTEST2 outcomes. Due to the complex sampling design (i.e., students nested within schools), two-level HLM was used to analyze the data where students (level 1) were nested within schools (level 2). This analytic approach was different from previous Globaloria studies (Chadwick & Gore, 2010, 2011). Previously, MANCOVA was used to examine the associations between Globaloria participation and student achievement outcomes. MANCOVA was a favorable approach when the study sample size was small; yet, this analytic method did not allow the researchers to take into account within- and between-school variations in student achievement outcomes. In other words, the MANCOVA assumes that each student's achievement is independent from others; however, in the real world, students from the same schools are more similar than students from different schools. Given the adequate sample size available for the Year 4 analyses, the more robust approach (e.g., HLM) provided the advantages of taking these within- and between-school variations in student achievement into account.

The preliminary examination of student demographic composition revealed that minority students (i.e., Black and Asian) each comprised less than 4% of the sample. Additionally, these two groups were significantly different in achievement outcomes from White students. These patterns suggest that it is more appropriate to examine the associations between participation and student achievement for each racial group independently. For the White student group, researchers used HLM, given that there was adequate sample size to do so. Due to the small sample size of the Black student group (valid $n < 40$), HLM was not appropriate; therefore, four one-way analyses of covariance (ANCOVAs) were conducted to examine the main effect of participation on student WESTEST2 outcomes by controlling for WESTEST2 pretest scores. Results showed that participating Black students and comparison Black students were not statistically different in WESTEST2 subtests (i.e., mathematics, science, reading, and social studies).⁸ Given the lack of significant findings, this population will not be discussed further in subsequent sections. Similarly,

⁸ This finding should be interpreted with caution given that the sample sizes were small.

the Asian student group had such a small sample size (valid $n < 10$) that it was not appropriate to perform any type of analyses on the group; therefore, this group will not be discussed further.

To perform the HLM analyses, student-level demographic information (e.g., student grade level, gender, income status) and pretest scores (e.g., 2010 WESTEST2 scale scores) were entered into the model to adjust for between-student variations in WESTEST2 achievement. Interaction terms were added into the HLM model to examine whether the relationships between Globaloria participation and student WESTEST2 outcomes differed by subgroups (e.g., gender, grade level, income status). Although not the main focus of this evaluation, three school-level contextual characteristics (e.g., percentage of low-income students, percentage of students scoring proficient and higher on the 2010 WESTEST2 math subscale, and percentage of students scoring proficient and higher on the 2010 WESTEST2 reading subscale) were entered into the model to adjust for between-school variations in student 2011 WESTEST2 outcomes. By including these school-level contextual factors into the model, researchers were able to understand whether school-level contextual factors (e.g., school performance measured by the percentage of students scoring proficient or higher) play a role in either “strengthening” or “weakening” the associations between Globaloria participation and student outcomes. For example, researchers examined whether Globaloria participation had a different impact on student achievement if the students were in schools with a high percentage of students scoring proficient or higher in math.

The relationships among school-level and individual-level factors can be quite complex; when there is no significant relationship between participation and student achievement, knowledge about the school-level characteristics may help researchers and program staff understand what may be influencing the outcomes (Lee, 2000). For program staff, this type of information is critical for the purpose of planning and programing. As such, studying contextual factors is a key concern for future research. In this particular study, adding these school-level contextual factors in the model was primarily exploratory, and the factors selected for inclusion were based on information that was readily available in the dataset. For future studies, Globaloria program staff are encouraged to collect some “proximal” school-level or teacher-level factors that may help to explain how Globaloria works to support student achievement. See Appendix D for level 1 and 2 HLM model equations, notations, and definitions of the variables.

Results

The main purpose of this report is to document the relationships between Globaloria participation and student achievement in the four subject domains assessed by the WESTEST2 (math, science, reading, and social studies). Due to insufficient sample sizes of the Asian and Black student populations, these two racial minority groups were removed from further analyses. The sections below only include data from the White student group. Tables 4-7 display detailed information from the HLM analyses. Additionally, Appendix D presents details about model building and technical reports of HLM analyses.

Mathematics

After controlling for school-level contextual factors (e.g., percent of students achieving reading proficiency on the 2010 WESTEST2, percent of students achieving math proficiency on the 2010 WESTEST2, school enrollment, and percent of low-income students) and student-level covariates (e.g., grade level, gender, income status), Globaloria participation was positively related to student math achievement, $\beta = 36.37$, $SE = 12.26$, $p = 0.004$, as shown in Table 4. Specifically, students who participated in Globaloria scored significantly higher than their comparison counterparts on the mathematics subtest. There were no significant differences between achievement and student grade level, gender, or low-income status. Additionally, there were also no significant differences by school-level contextual factors (e.g., percent of students achieving reading proficiency on the 2010 WESTEST2, percent of students achieving math proficiency on the 2010 WESTEST2, school enrollment, and percent of low-income students). Overall, this HLM model explained 38% of within-school variations and almost 100% of between-school variations in student math outcomes.

Table 4. Results of Two-Level HLM: Mathematics

Parameter Estimates	Null Model		Final Model	
	β	SE	β	SE
Mean of Math Intercept (β_0)	658.00***	2.68	623.93***	9.15
High Poverty School			2.89	5.74
% 2010 Math Proficiency			0.61	0.32
% 2010 Reading Proficiency			-0.04	0.35
Mean of Participation Slope (β_1)			36.37**	12.26
High Poverty School			-7.66	6.06
% 2010 Math Proficiency			-0.32	0.36
% 2010 Reading Proficiency			-0.26	0.39
Mean of Grade Slope (β_2)			12.94*	5.47
Mean of Gender Slope (β_3)			5.24*	2.51
Mean of Low-Income Slope (β_4)			-29.42	15.16
High Poverty School			11.72*	5.20
% 2010 Math Proficiency			-0.24	0.35
% 2010 Reading Proficiency			0.70	0.40
Mean of Math @ 2010 Slope (β_8)			0.37***	0.04
Mean of Reading @ 2010 Slope (β_9)			0.13**	0.04

Parameter Estimates	Null Model		Final Model	
	β	SE	β	SE
Mean of Science @ 2010 Slope (β_{10})			0.11**	0.03
Mean of Social Studies @ 2010 Slope (β_{11})			0.18**	0.06
Participation X Grade (β_5)			-11.17	6.43
Participation X Gender (β_6)			-0.64	3.95
Participation X Low-Income (β_7)			-5.06	4.77
Variance Components				
Between school variation (var. of u_{0j} , τ_{00})	282.45***		0.93	
Within school variation (var. of γ_{ij} , σ_{00})	1870.41		1161.68	
No. of parameters	3		23	
Deviance (FIML)	10324.53		785.52	
Intra-Class Correlation Coefficient (ICC)	13.1%		0.1%	
Within school variation explained (level-1 R^2)			37.9%	
Between school variation explained (level-2 R^2)			99.7%	

Note. Robust estimations are reported. Missing values are handled by using Full Information Maximum Likelihood (FIML). The calculation of R^2 is based on Raudenbush & Bryk's version (2002).

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

Science

After controlling for school-level contextual factors (e.g., percent of students achieving reading proficiency on the 2010 WESTEST2, percent of students achieving math proficiency on the 2010 WESTEST2, school enrollment, and percent of low-income students) and student-level covariates (e.g., grade level, gender, income status), there was no significant difference between participating and comparison students on science achievement, $\beta = 5.52$, $SE = 20.33$, $p = 0.786$. Specifically, students who participated in Globaloria did not score higher or lower than nonparticipating students as they scored about the same. Additionally, there were no significant differences among subgroups (e.g., grade level, gender, low-income status). Overall, this HLM model explained 33% of within-school variations and 85% of between-school variations in student science achievement. Results of the HLM analysis for science outcomes are reported in Table 5.

Table 5. Results of Two-Level HLM: Science

Parameter Estimates	Null Model		Final Model	
	β	SE	β	SE
Mean of Science Intercept (β_0)	618.12***	3.99	611.69***	16.82
High Poverty School			-11.01	7.80
% 2010 Math Proficiency			1.38**	0.47
% 2010 Reading Proficiency			-1.24*	0.49
Mean of Participation Slope (β_1)			5.52	20.33
High Poverty School			1.44	10.95
% 2010 Math Proficiency			-2.35***	0.57
% 2010 Reading Proficiency			1.87**	0.60

Parameter Estimates	Null Model		Final Model	
	β	SE	β	SE
Mean of Grade Slope (β_2)			22.34**	7.38
Mean of Gender Slope (β_3)			-2.09	4.74
Mean of Low-Income Slope (β_4)			-5.64	21.38
High Poverty School			14.54**	5.06
% 2010 Math Proficiency			-0.53	0.35
% 2010 Reading Proficiency			0.44	0.49
Mean of Math @ 2010 Slope (β_8)			0.11**	0.04
Mean of Reading @ 2010 Slope (β_9)			0.21***	0.06
Mean of Science @2010 Slope (β_{10})			0.27***	0.05
Mean of Social Studies @ 2010 Slope (β_{11})			0.25**	0.08
Participation X Grade (β_5)			5.43	10.14
Participation X Gender (β_6)			4.65	6.51
Participation X Low-Income (β_7)			-4.77	5.24
Variance Components				
Between school variation (var. of u_{0j} , τ_{00})	733.7***		84.69***	
Within school variation (var. of γ_{ij} , σ_{00})	1652.6		1117.44	
No. of parameters	3		23	
Deviance (FIML)	6593.5		6282.22	
Intra-Class Correlation Coefficient (ICC)	30.75%		7.0%	
Within school variation explained (level-1 R^2)			32.7%	
Between school variation explained (level-2 R^2)			85.1%	

Note. Robust estimations are reported. Missing values are handled by using FIML. The calculation of R^2 is based on Raudenbush & Bryk's version (2002).

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

As stated previously, the relationship between Globaloria participation and student science outcomes was not statistically significant; however, a significant relationship was found between participation and school-level math proficiency (i.e., percentage of students scoring proficient or higher in math), $\beta = -2.35$, $SE = 0.57$, $p < 0.001$. As shown in Figure 2, participating students scored higher on the science subtest than their comparison counterparts if they attended a school with lower percentages of students scoring proficiency in math. In contrast, comparison students did better in science than participating students if they attended a school with higher percentages of students scoring proficiency in math. This finding suggests that, although the relationship between participation and science achievement was not significant *overall*, the relationship between Globaloria participation and student science achievement trended toward a stronger positive relationship in low math-performing schools; this relationship seems to be negated in schools with stronger math achievement. Although outside of the scope of this study, more research is needed to further understand this finding.

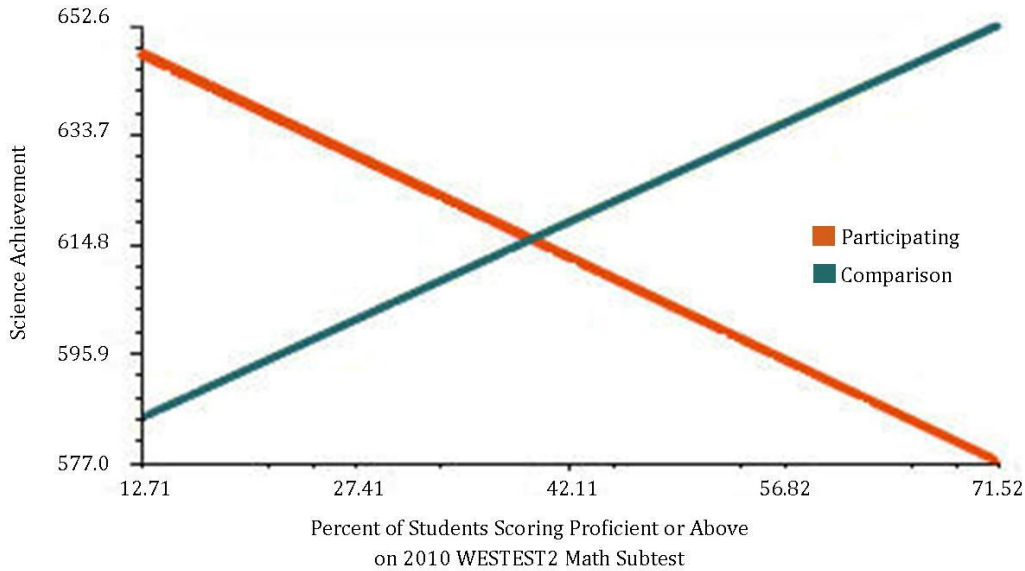


Figure 2. Relationship between participation and math proficiency on science subtest.

Additionally, a significant relationship was found between participation and school-level reading proficiency (i.e., percentage of students scoring proficient or higher in reading). This relationship is shown in Figure 3. Participating students scored significantly higher on the science subtest than their comparison counterparts if they attended schools with a higher percentage of students scoring proficiency in reading, $\beta = 1.87$, $SE = 0.60$, $p = 0.003$. This finding suggests that a school with a generally higher reading achievement level may provide a stronger foundation for Globaloria's effect on student science achievement than does a school without as high a percentage of students scoring proficiency in reading. It is not clear why the comparison students appeared to perform less well in science if they were in a school with generally high rates of reading proficiency, nor is it clear why the relationship with Globaloria participation is reversed depending on whether the school performs well in reading versus math; these relationships are worthy of further investigation.

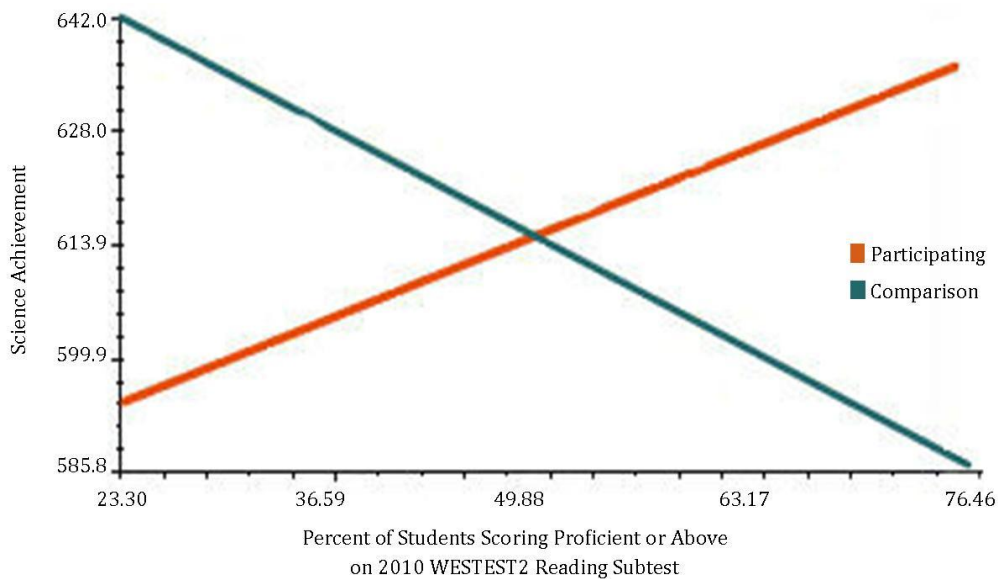


Figure 3. Relationship between participation and reading proficiency on science subtest.

Reading

As shown in Table 6, after controlling for school-level contextual factors (e.g., percent of students achieving reading proficiency on the 2010 WESTEST2, percent of students achieving math proficiency on the 2010 WESTEST2, school enrollment, and percent of low-income students) and student-level covariates (e.g., grade level, gender, income status), Globaloria participation was positively related to student reading achievement, $\beta = 36.25$, $SE = 13.57$, $p = 0.008$. However, no significant differences were found among student-level variables (e.g., grade level, gender, low-income status). Overall, this HLM model explained 53% of within-school variations and 97% of between-school variations in student science achievement.

Table 6. Results of Two-Level HLM: Reading

Parameter Estimates	Null Model		Final Model	
	β	SE	β	SE
Mean of Reading Intercept (β_0)	493.98***		476.95***	9.77
High Poverty School			-3.03	4.95
% 2010 Math Proficiency			-0.20	0.27
% 2010 Reading Proficiency			0.87**	0.23
Mean of Participation Slope (β_1)			36.25**	13.57
High Poverty School			-1.51	7.00
%2010 Math Proficiency			0.09*	0.39
% 2010 Reading Proficiency			0.05	0.39
Mean of Grade Slope (β_2)			-6.26	5.66
Mean of Gender Slope (β_3)			-17.07***	3.02
Mean of Low-Income Slope (β_4)			-9.32	10.18
High Poverty School			3.85	4.23
% 2010 Math Proficiency			0.00	0.39
% 2010 Reading Proficiency			0.02	0.45
Mean of Math @ 2010 Slope (β_8)			0.10*	0.04
Mean of Reading @ 2010 Slope (β_9)			0.54***	0.08
Mean of Science @ 2010 Slope (β_{10})			0.001	0.04
Mean of Social Studies @ 2010 Slope (β_{11})			0.38***	0.08
Participation X Grade (β_5)			-1.37	7.31
Participation X Gender (β_6)			7.69	4.59
Participation X Low-Income (β_7)			5.99	4.25
Variance Components				
Between school variation (var. of u_{0j} , τ_{00})	278.55***		8.97	
Within school variation (var. of y_{ij} , σ_{00})	2347.74		1098.99	
No. of parameters	3		23	
Deviance (FIML)	10127.30		9355.05	
Intra-Class Correlation Coefficient (ICC)	10.6%		0.8%	

Parameter Estimates	Null Model		Final Model	
	β	SE	β	SE
Within school variation explained (level-1 R^2)			53.2%	
Between school variation explained (level-2 R^2)			96.8%	

Note. Robust estimations are reported. Missing values are handled by using FIML. The calculation of R^2 is based on Raudenbush & Bryk's version (2002).

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

Although participation in Globaloria seems to have a positive effect on student reading achievement, further analyses revealed a significant relationship between participation and student reading achievement when school-level math proficiency (i.e., percentage of students scoring proficient or higher in math) was taken into consideration, $\beta = 0.09$, $SE = 0.39$, $p = 0.019$. Figure 4 depicts that participating students outperformed their comparison counterparts when they both came from schools with a lower percentage of students scoring proficiency or higher in math. In contrast, comparison students outperformed the participating students when they both came from schools with a higher percentage of students scoring proficient or higher in math. This suggests that students from lower math-achieving schools seem to benefit more from Globaloria in terms of reading achievement. Although outside of the scope of this study, more research is needed to further understand this finding.

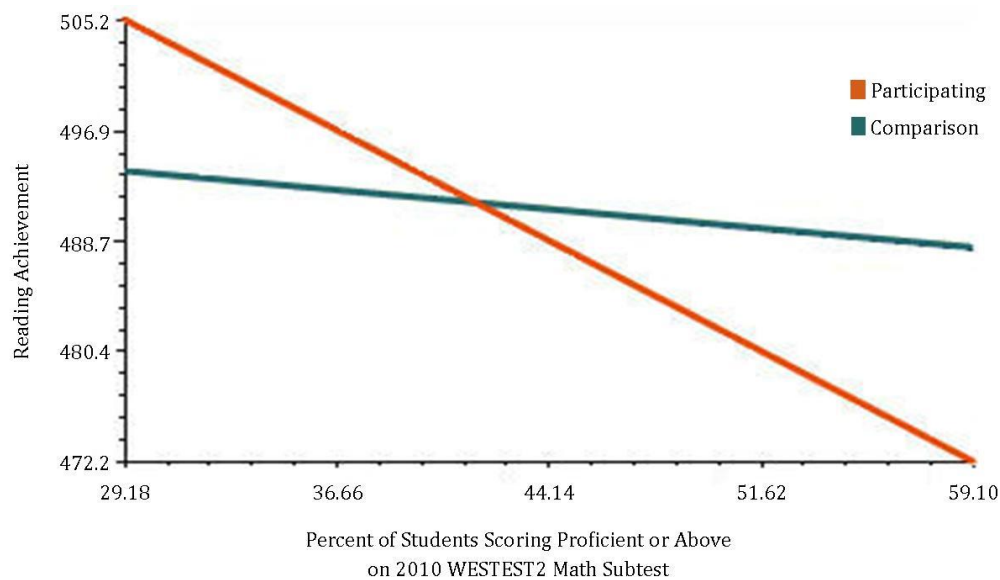


Figure 4. Relationship between participation and math proficiency on reading subtest.

Social Studies

After controlling for school-level contextual factors (e.g., percent of students achieving reading proficiency on the 2010 WESTEST2, percent of students achieving math proficiency on the 2010 WESTEST2, school enrollment, and percent of low-income students) and student-level covariates (e.g., grade level, gender, income status), there was no significant difference between participating and comparison students on social studies achievement, $\beta = 9.43$, $SE = 9.11$, $p = 0.301$. Specifically, students who participated in Globaloria did not score higher or lower than

nonparticipating students; they scored about the same. Additionally, there were no significant differences by school-level factors or student-level covariates. Results of the HLM are displayed in Table 7. This HLM model accounted for 48% of within-school variations and almost 100% of between-school variations in student social studies outcomes.

Table 7. Results of Two-Level HLM: Social Studies

Parameter Estimates	Null Model		Final Model	
	β	SE	β	SE
Mean of Social Studies Intercept (β_0)	411.43***	1.52	400.21***	9.26
High Poverty School			2.9	3.81
% 2010 Math Proficiency			-0.03	0.19
% 2010 Reading Proficiency			0.33	0.28
Mean of Participation Slope (β_1)			9.43	9.11
High Poverty School			1.25	3.61
% 2010 Math Proficiency			-0.36	0.21
% 2010 Reading Proficiency			0.02	0.28
Mean of Grade Slope (β_2)			-6.77*	3.21
Mean of Gender Slope (β_3)			2.3	2.97
Mean of Low-Income Slope (β_4)			3.9	9.04
High Poverty School			-3.9	3.64
% 2010 Math Proficiency			0.18	0.21
% 2010 Reading Proficiency			-0.24	0.26
Mean of Math @ 2010 Slope (β_8)			0.12***	0.03
Mean of Reading @ 2010 Slope (β_9)			0.21***	0.04
Mean of Science @ 2010 Slope (β_{10})			0.05	0.03
Mean of Social Studies @2010 Slope (β_{11})			0.34***	0.04
Participation X Grade (β_5)			1.06	3.57
Participation X Gender (β_6)			2.98	4.32
Participation X Low-Income (β_7)			1.11	3.24
Variance Components				
Between school variation (var. of u_{0j} , τ_{00})	62.41***		0.20	
Within school variation (var. of γ_{ij} , σ_{00})	987.12		514.42	
No. of parameters	3		23	
Deviance (FIML)	9350.79		8688.89	
Intra-Class Correlation Coefficient (ICC)	6.0%		0.0%	
Within school variation explained (level-1 R^2)			47.9%	
Between school variation explained (level-2 R^2)			99.7%	

Note. Robust estimations are reported. Missing values are handled by using FIML. The calculation of R^2 is based on Raudenbush & Bryk's version (2002).

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

Key Cross-Year Findings

While it is difficult to compare findings across the three studies, the similarities and differences are discussed to the extent possible. In the pilot study (Chadwick & Gore, 2010), Globaloria participants outperformed their comparison counterparts in science and social studies achievement; however, in the replication study (Chadwick & Gore, 2011), participating students outperformed their comparison counterparts in science. The change in findings was likely due to the increase sample size and improved matching procedures (using PSM instead of Euclidean distances).

In the current study, Globaloria participants outperformed their comparison counterparts in mathematics and reading, a finding that is different from the findings of the previous two studies. It is likely that this is a result of the within- and between-school variations in student achievement that are taken into consideration in HLM, but are not taken into consideration in the MANCOVA used in the pilot (Chadwick & Gore, 2010) and replication studies (Chadwick & Gore, 2011). The increased sample size in the current study may also have played a role in the different findings. In previous studies, it is a possibility that the sample size was too small to detect the difference in reading achievement. It is expected that Globaloria participation would increase reading achievement because reading comprehension and writing is key to the Globaloria program. This study will be replicated in Year 5 using the same design to gain further evidence of Globaloria participation on mathematics and reading achievement.

Conclusions and Recommendations

This study examined the relationships between Globaloria participation and student achievement, as measured by the WESTEST2. Edvantia researchers conducted a series of HLM analyses for each of the four core subject areas, including mathematics, science, reading, and social studies. Results showed that Globaloria participation was positively associated with math and reading achievement. Overall, relationships between participation and achievement were not statistically significant in the subject areas of science and social studies. Further investigation suggested that the lack of significant relationships on certain outcomes may be explained by school-level contextual factors (e.g., percent of students achieving reading proficiency on the 2010 WESTEST2, percent of students achieving math proficiency on the 2010 WESTEST2, school enrollment, and percent of low-income students). In other words, the effects of Globaloria participation may be dependent on the characteristics of schools that students attended. The following bullets summarize key findings by core subject area:

- Students who participated in Globaloria scored significantly higher than nonparticipating students on the **mathematics** subtest.
- While there were no significant differences between participating and nonparticipating students' **science** scores overall, the influence of Globaloria participation on science scores seemed to be affected by the percentage of students achieving proficiency in mathematics at the school. Specifically, participating students in schools with low mathematics proficiency scored significantly higher on the science subtest than their nonparticipating counterparts.
- Students who participated in Globaloria scored significantly higher than nonparticipating students on the **reading** subtest. Similar to the finding for science, the effects of Globaloria participation were stronger in schools with a low percentage of students achieving mathematics proficiency.
- No significant differences were found between the scores of participants and nonparticipants on the **social studies** subtest.

Cross-Year Comparison

As mentioned, the design of the current study has been greatly improved over the pilot and replication studies. The first major difference among the three studies was the increase in sample size each year. As the sample size increases, the power of the statistical test increases, meaning that a better chance exists that a relationship among variables will be detected if it actually exists.

The second difference among the studies was the matching procedure used. In the pilot study, Edvantia used Euclidean distances to match participants; in the replication and current studies, PSM was used. In the field of education and evaluation where random assignment is not always feasible or ethical, PSM provides advantages to examine the effects of interventions such as Globaloria. The advantage of using the propensity score is that it combines information from all the other covariates into a single variable. The literature indicates that if matching variables are normally distributed and independent, then Euclidean distance may be a good method for matching (Rubin, 2006). However, it is not likely that the matching variables are independent. For example, students who score high on one subtest are likely to score high on the other subtests. This means that the mathematics and reading variables are not independent; they are related. In this vein, the use of PSM is the recommended method in the literature (Rubin, 2006).

The third difference was the research and analyses design. Previously, MANCOVA was used to examine the associations between Globaloria participation and student achievement outcomes. MANCOVA was a favorable approach when the study sample size was small; yet, this analytic method did not allow the researchers to take into account within- and between-school variations in student achievement outcomes. In other words, the MANCOVA assumes that each student's achievement is independent from others; however, in the real world, students from the same schools are more similar than students from different schools. Given the adequate sample size available for the Year 4 analyses, the more robust approach (e.g., HLM) provided the advantages of taking these within- and between-school variations in student achievement into account.

While it is difficult to compare the findings across the three studies, researchers compared the key findings to the extent possible. In the pilot study, Globaloria participants outperformed their comparison counterparts in science and social studies achievement; however, only the difference for science was upheld in the replication study. In the current study, Globaloria participants outperformed their comparison counterparts in mathematics and reading, which is different from the findings of the previous two studies. It is likely that the change in findings is due to the increased sample size, better matching procedures, and a more robust design that takes within- and between-school variations in student achievement into consideration. It is also possible that the focus of the games that were created played a role in the shift in findings. This is worthy of further investigation.

Recommendations

Based on these findings, researchers offer several recommendations for future research on Globaloria:

- The Workshop is encouraged to re-examine the logic model suggested by Edvantia in 2008 (Knestis, 2008) to identify “mediators” to help explain the processes that link program participation with student outcomes. These mediators are key for future program scale-up.
- The Workshop program staff are encouraged to consider replicating the study methodology using a more diverse sample of students than in the current study. This likely would require drawing upon Globaloria implementation sites other than West Virginia, a rather demographically homogenous state.
- Researchers interested in studying Globaloria are encouraged to consider a longitudinal study with middle school students to investigate the extent to which Globaloria participation may have long-term effects on students, both in terms of achievement outcomes and college- and career-readiness outcomes.
- It is recommended that future studies bring together research that has been conducted on teacher characteristics and behaviors (i.e., proximal moderators), and the effects of Globaloria participation on student achievement outcomes.
- Findings of this study suggest that “school context” matters. It is recommended that future research include examination of factors that are malleable (e.g., educator practice and interaction with students) so these factors can be incorporated into the Globaloria model as it evolves.
- It is important that future studies investigate “self-selection” factors that may be playing a part in the evidenced relationships between participation and achievement. Most students who participate in Globaloria are likely to be on track to graduate; if they were not, they would not have room in their course schedule for an elective on game design. They may have stronger interests in mathematics, technology, and science than

students who do not participate in Globaloria, and the participating students may be more motivated to learn new things than their nonparticipating counterparts. Additionally, assessing students' self-selection factors will increase the power of quasi-experimental designs (Luellen, Shadish, & Clark, 2005; Stuart & Rubin, 2007), and help researchers make valid attributions regarding the effects of the program versus the traits of the students.

- Future studies should examine the relationship between the type of game (e.g., topic/content) and students' performance on the related area of the WESTEST2. This information is readily available as all games are available on the Globaloria platform and have been reviewed and coded for research by Dr. Rebecca Reynolds at Rutgers University, a primary research partner for the Globaloria project.

Limitations

Readers should be aware of the limitations of this study. First, although PSM provides an advantage by creating a comparable counterfactual group by using various covariates, it has its own limitations. A specific limitation is that similarity between the Globaloria group and comparison group can only be determined within the covariates used in the matching process. Researchers do not know whether these two groups were different in other aspects. In addition, PSM relies on the reliability of the covariates (Luellen et al., 2005; Stuart & Rubin, 2007). As Luellen and colleagues (2005) stated,

Even in cases in which there are no important pretreatment differences between groups on observed covariates, we cannot be sure that the same is true for unobserved covariates—and we have no reason to assume that scores on the unobserved covariates are randomly distributed across groups, as we do with randomized experiments (p. 546-547).

Second, lacking self-selection indicators may minimize the generalizability of this study. For example, it is unclear who, why, and how students decide to participate in Globaloria. If students who were more motivated to learn were more likely to participate and remain in the program, the effects of Globaloria were likely to be confounded with this self-selection factor. Also, self-selection factors are the key covariates that should be included when conducting PSM; however, these factors were not available for this study. Future study including these variables may increase the power of the design as well as results.

A third limitation is related to the cross-year comparison. Given that the samples are different from year to year and the matching procedure and research design changed overtime, it is difficult to draw conclusions across years about the effects of Globaloria. This study will be replicated in Year 5 using the same design to examine whether the mathematics and reading findings are supported.

References

- Austin, P. C. (2011). A tutorial and case study in propensity score analysis: An application to estimating the effect of in-hospital smoking cessation counseling on mortality. *Multivariate Behavioral Research*, 46(1), 119-151.
- Center for Improving Methods for Quantitative Policy Research. (2012). Propensity score matching [Online video]. Available from <http://www.ipr.northwestern.edu/qeworkshops/stage2/index.html>
- Chadwick, K., & Gore, J. N. (2010). *Globaloria pilot study: The relationship of Globaloria participation and student achievement*. Charleston, WV: Edvantia, Inc.
- Chadwick, K., & Gore, J. N. (2011). *Globaloria replication study: Examining the robustness of relationships between Globaloria participation and student achievement*. Charleston, WV: Edvantia, Inc.
- Edvantia, Inc. (2008). *A chronicle of West Virginia's 21st century learning initiative*. Charleston, WV: Author.
- Harder, V. S., Stuart, E. A., & Anthony, J. C. (2010). Propensity score techniques and the assessment of measured covariate balance to test causal associations in psychological research. *Psychological Methods*, 15(3), 234-249.
- Harel, C. I. (1991). *Children designers: Interdisciplinary constructions for learning and knowing mathematics in a computer-rich school*. Unpublished manuscript, Westport, CT.
- Harel, C. I., Oliver, A., & Sullivan, S. (2010). *A blueprint for global learning and leadership in the 21st century. Transforming education through the Globaloria learning network (2009-2010 Annual Report)*. New York, NY: World Wide Workshop Foundation.
- Hill, J. (2004). *Reducing bias in treatment effect estimation in observational studies suffering from missing data*. Unpublished manuscript, Columbia University Institute for Social and Economic Research and Policy (ISERP).
- Knestis, K. (2008). *Understanding Globaloria as the subject of research: An agenda for future study*. Charleston, WV: Edvantia, Inc.
- Lee, V. E. (2000). Using hierarchical linear modeling to study social contexts: the case of school effects. *Educational Psychologist*, 35(2), 125-141.
- Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review*, 29(6), 530-558.
- Nicholson, B., Alley, R., Green, J., & Lawson, D. (2009). *An analysis of the effects of a technology program on students' academic performance: Are these Vygotsky's children?* Retrieved from <http://www.worldwideworkshop.org/pdfs/Globaloria-StudentAcademicPerformance.pdf>
- Parsons, L. S. (2001, April). *Reducing bias in a propensity score matched-pair sample using greedy matching techniques* [Paper 214-26.]. Paper presented at the SAS® Users Group International Conference 26, Long Beach, California.

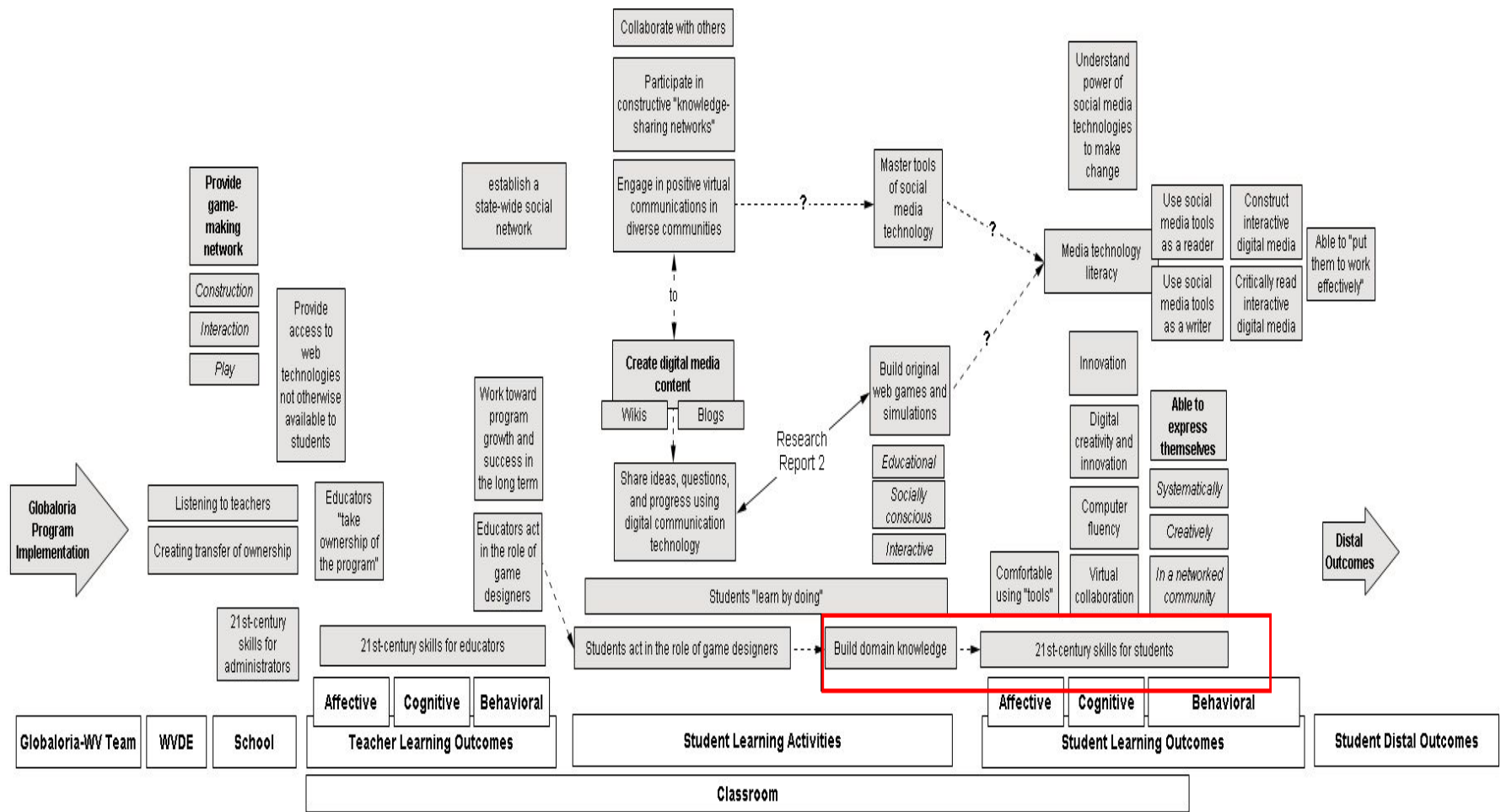
- Raudenbush, S. W., & Bryk, A. S. (Eds.). (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage Publications.
- Reynolds, R., & Chiu, M. (2012). *Contribution of motivational orientations to student outcomes in a discovery-based program of game design learning*. A paper presented at the International Conference of the Learning Sciences, Sydney, Australia.
- Reynolds, R., & Kim, G. (2012). *The contribution of educator factors to student outcomes in a discovery-based program of game design learning*. A paper presented at the annual conference of the Association of Library and Information Science Education (ALISE).
- Reynolds, R., Scialdone, M., & Harel, C. I. (2010). *Evidence of high school students' development of contemporary learning abilities in a game design program in rural West Virginia*. New York, NY: World Wide Workshop Foundation.
- Rogers, P. J., Pertosino, A., Huebner, T. A., & Hacs, T. A. (2000). *Program theory evaluation: Practice, promise, and problems*. San Francisco, CA: Jossey-Bass.
- Rubin, D. B. (2006). *Matching sampling for causal effects*. Cambridge, UK: Cambridge University Press.
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services & Outcomes Research Methodology*, 2, 169-188.
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, 25, 1-21.
- Stuart, E. A., & Rubin, D. B. (2007). Best practices in quasi-experimental designs: Matching methods for causal inference. In J. Osborne (Ed.), *Best Practices in Quantitative Social Science* (Vol. 11, pp. 155-176). Thousand Oaks, CA: Sage Publications.
- Whitehouse, P. (2009). *Networked teacher professional development: The case of Globaloria*. New York, NY: World Wide Workshop Foundation.
- World Wide Workshop Foundation. (2008). *The case for the Globaloria network in West Virginia: Empowering West Virginia youth to create and collaborate online with a 21st-century game-making curriculum*. New York, NY: World Wide Workshop Foundation.
- World Wide Workshop Foundation. (2009). *Globaloria: Empowering students and educators with 21st-century digital literacy through a game-making learning network*. New York, NY: World Wide Workshop Foundation.
- World Wide Workshop Foundation. (2012). *World Wide Workshop: Social media technology & learning*. New York, NY: World Wide Workshop Foundation.

Appendices

Appendix A

School-Level Globaloria Implementation and Learning Outcomes Logic Model

Appendix A: School-Level Globaloria Implementation and Learning Outcomes Logic Model



Appendix B

Participating and Matched Comparison School Selection Criteria

Appendix B: Participating and Matched Comparison School Selection Criteria

Match	School	County	% of Students Passing Reading Proficiency	% of Students Passing Math Proficiency	School Enrollment	% of Low-Income Students
Participating Schools						
1	Braxton County High School	Braxton	37.94	39.06	658	48.6
2	Buffalo High School	Putnam	53.05	61.21	274	44.9
3	Cameron High School	Marshall	44.59	41.46	326	46.6
4	Capital High School	Kanawha	48.47	34.57	1,236	48.2
5	Chapmanville Middle School	Logan	45.50	35.27	574	56.4
6	Chapmanville Regional High School	Logan	39.29	29.12	712	48.9
7	Doddridge County High School	Doddridge	47.08	34.43	374	48.4
8	Dunbar Middle School	Kanawha	41.13	33.33	440	64.5
9	Eastern Greenbrier Middle School	Greenbrier	48.58	43.59	865	51.0
10	Fayetteville High School	Fayette	38.79	35.41	510	52.0
11	George Washington High School	Kanawha	75.52	66.58	1,115	17.9
12	Greenbrier East High School	Greenbrier	51.09	46.60	1,129	49.2
13	Greenbrier West High School	Greenbrier	43.32	31.27	430	60.9
14	Hurricane High School	Putnam	61.41	63.98	1,142	24.7
15	Kasson Middle School	Barbour	39.53	46.51	200	45.0
16	Liberty High School	Raleigh	63.65	57.21	532	46.1
17	Logan Middle School	Logan	46.55	37.70	796	62.1
18	Logan Senior High School	Logan	34.98	29.20	826	41.9
19	Man High School	Logan	46.95	31.30	478	61.3
20	Man Middle School	Logan	41.63	29.18	392	51.5
21	Oak Glen High School	Hancock	47.19	52.24	616	41.2
22	Oak Hill High School	Fayette	40.37	33.33	771	49.7

Match	School	County	% of Students Passing Reading Proficiency	% of Students Passing Math Proficiency	School Enrollment	% of Low-Income Students
23	Philip Barbour High School	Barbour	34.21	37.40	773	54.9
24	Riverside High School	Kanawha	35.82	27.36	1,244	49.3
25	Sandy River Middle School	McDowell	41.57	49.10	272	72.1
26	South Charleston Middle School	Kanawha	63.01	53.31	414	49.8
27	South Harrison High School	Harrison	41.77	37.97	422	39.6
28	Spring Valley High School	Wayne	49.46	39.60	1,041	34.4
29	Tygarts Valley High School	Randolph	37.06	25.80	478	63.2
30	Webster County High School	Webster	36.76	32.94	483	56.7
31	Weir High School	Hancock	45.53	48.27	619	37.5
32	Western Greenbrier Middle School	Greenbrier	41.58	40.31	319	66.5
33	Wheeling Park High School	Ohio	51.65	44.25	1,690	42.3
PARTICIPATING STUDENT SAMPLE MEAN			45.91	40.87	671	49.3

Match	School	County	% of Students Passing Reading Proficiency	% of Students Passing Math Proficiency	School Enrollment	% of Low-Income Students
Comparison Schools						
1	Mount Hope High School	Fayette	23.30	12.71	302	76.8
2	Midland Trail High School	Fayette	29.46	29.04	331	62.2
3	Paw Paw High School	Morgan	40.25	33.76	99	57.6
4	Bridgeport High School	Harrison	76.46	71.52	794	13.4
5	Short Line School	Wetzel	40.07	32.62	453	62.5
6	Paden City High School	Wetzel	59.20	52.00	153	48.4
7	Saint Albans High School	Kanawha	46.99	41.01	1,070	38.0
8	East Bank Middle School	Kanawha	36.19	31.86	437	63.8
9	Oakvale School	Mercer	48.85	46.56	204	72.5
10	John Marshall High School	Marshall	36.76	37.65	1,220	39.7
11	Parkersburg South High School	Wood	55.89	42.16	1,580	39.6
12	Martinsburg High School	Berkeley	44.58	38.48	1,685	50.0
13	Morgantown High School	Monongalia	59.60	61.93	1,616	32.5
14	St Marys High School	Pleasants	46.51	36.07	425	36.0
15	Paw Paw Junior High School	Morgan	40.25	33.76	99	57.6
16	Herbert Hoover High School	Kanawha	48.97	39.29	739	37.5
17	Clay Middle School	Clay	48.83	43.15	449	70.8
18	Winfield High School	Putnam	57.96	61.92	823	19.4
19	Huntington Middle School	Cabell	46.94	34.75	612	55.6
20	Shady Spring High	Raleigh	57.45	57.02	809	30.9
21	Pikeview High School	Mercer	37.08	39.72	708	51.4
22	Washington High School	Jefferson	47.82	45.93	1,110	22.4
23	Wirt County High School	Wirt	44.11	46.21	335	49.9
24	Point Pleasant High School	Mason	45.13	40.55	1,214	54.2

Match	School	County	% of Students Passing Reading Proficiency	% of Students Passing Math Proficiency	School Enrollment	% of Low-Income Students
25	Van Devender Junior High School	Wood	39.51	44.35	373	69.4
26	Enslow Middle School	Cabell	30.06	17.13	284	72.2
27	Matewan High School	Mingo	41.40	35.43	196	70.9
28	Hedgesville High School	Berkeley	46.88	38.50	1,691	34.8
29	Independence High School	Raleigh	42.73	36.16	687	45.3
30	Van Junior/Senior High School	Boone	40.66	35.07	249	46.6
31	Brooke High School	Brooke	42.76	47.61	1,108	37.1
32	Glenwood School	Mercer	50.28	39.92	733	52.1
33	Nitro High School	Kanawha	64.36	59.10	755	28.9
COMPARISON STUDENT SAMPLE MEAN			45.98	41.30	707	48.5

Appendix C

Propensity Score Matching Steps and Procedures

Appendix C: Propensity Score Matching Steps and Procedures

In the field of education and evaluation research, there is an increasing use of propensity score matching (PSM) methods to adjust a treatment effect based on the function of observed variables (i.e., covariates) in nonrandomized observational studies (Stuart, 2010). PSM provides advantages in educational research where random assignment is not always feasible or ethical. In general, PSM consists of several analytic steps, including selecting covariates, estimating propensity score for matching, and diagnosing the matches (Stuart, 2010; Stuart & Rubin, 2007). Each step conducted for this report is described briefly.

Selection of Covariates

The logic behind PSM is that, in the absence of an experimental design, assignment to treatment is frequently nonrandom. Hence, units (e.g., schools, students, teachers) receiving treatment and those without treatment may differ in many ways that may affect both participation (i.e., self-selection factors) and the outcome of interest (i.e., pre-test scores, demographic attributes). To avoid the biased estimation of a treatment effect, PSM selects a comparison group that is similar to a treatment group based on observed covariates. Therefore, the main goal of PSM is to achieve balance on observed covariates through careful matching on the propensity score that is a function of observed covariates. In this sense, the selection of observed covariates is critical. In practice, three types of covariates commonly used in PSM are self-selection variables (e.g., individual preference, individual motivation), individual demographic information (e.g., gender, race, income status), and individual pre-test scores (e.g., pre-test achievement data) (Center for Improving Methods for Quantitative Policy Research, 2012; Stuart, 2010).

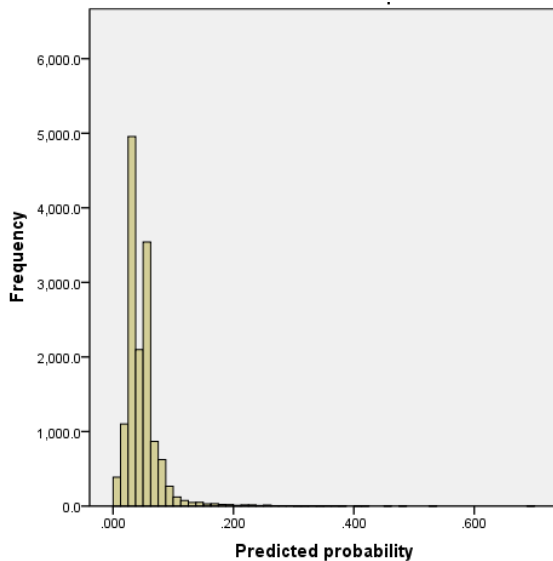
In this evaluation study, individual demographic information and pretest scores were used to estimate propensity scores for each individual, including gender, race, low-income student status, Limited English Proficiency (LEP) status, disability status, grade level, and four WESTEST2 subtests scores. In addition, the matching involved two stages. The first stage was to select a group of comparison schools matching to the group of Globaloria participants. Four school-level covariates were used to select the matched comparison schools: school enrollment, percentage of low-income students, percentage of students achieving proficiency on the 2010 WESTEST2 reading subtest, and percentage of students achieving proficiency on the 2010 WESTEST2 math subtest. The second stage was to select a group of comparison students from the identified comparison schools matching to the group of Globaloria participating students. This two-stage procedure ensures that the selected comparison schools were similar to the Globaloria participating schools, and students who were selected from the comparison schools were similar to the Globaloria participants prior to their participation in Globaloria.

Estimation of Propensity Scores

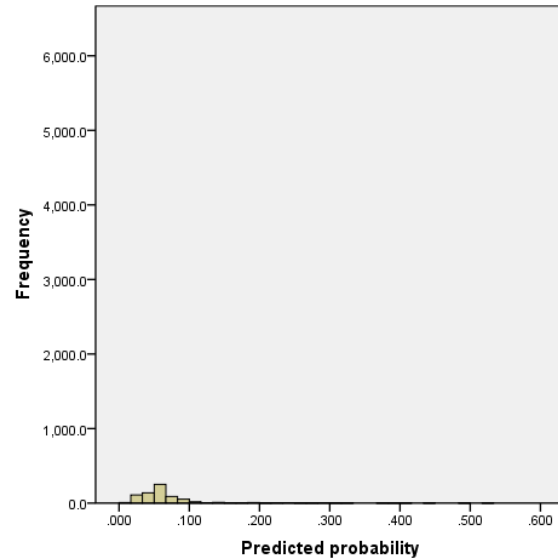
The next step was to conduct a logistic regression to obtain a measure describing the “distance” between two individuals, so called propensity score. First, logistic regression with school-level data was conducted to select the matched comparison schools, followed by student-level matching. Both procedures used one-on-one greedy matching algorithms that allows a comparison case to be selected only once (Parsons, 2001).

Since PSM methods require full information to create propensity scores, multiple imputations were conducted to impute missing values (Hill, 2004) before conducting logistic regression. In this procedure, individual demographic information were used as predictors but

were not imputed, whereas four individual subtests of 2010 WESTEST2 scores (i.e., the pretest) were used as predictors and were imputed. Once missing values were imputed, the next step was to obtain propensity scores by entering the 11 identified covariates (as aforementioned) and 16 interaction terms into the logistic regression model. The final sample included 701 Globaloria participants and 701 matched comparison students. Figure C1 shows the distributions of propensity scores before the matching. Figure C2 shows the distributions of propensity scores by groups after the matching.

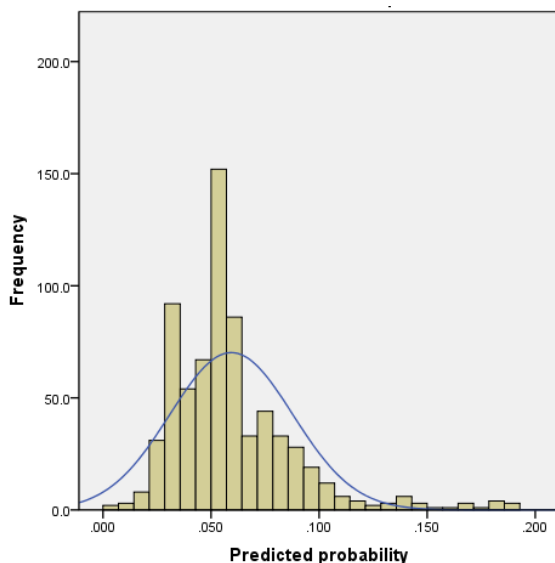


**Distribution of Propensity Score Before the Matching:
Comparison Group**

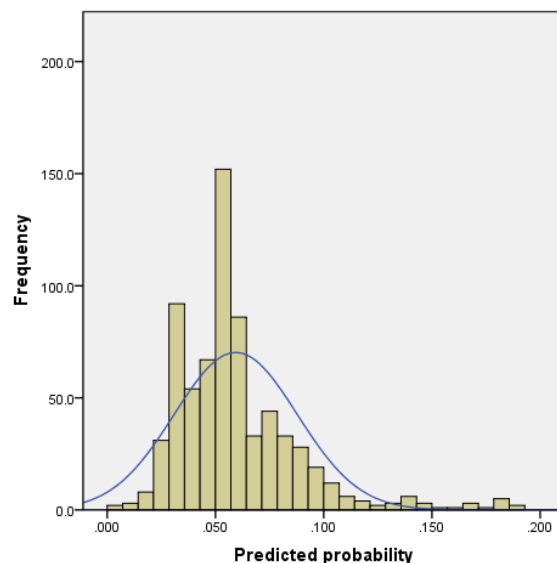


**Distribution of Propensity Score Before the Matching:
Participating Group**

Figure C1. Distributions of propensity scores by groups before the matching.



**Distribution of Propensity Score After the Matching:
Comparison Group**



**Distribution of Propensity Score After the Matching:
Participating Group**

Figure C2. Distributions of propensity scores by groups after the matching.

Diagnosis of the Matches

Once the comparison students were selected, the final step was to check for the covariate balance in the matched groups, which is defined as “the similarity of the empirical distributions of the full set of covariates in the matched participant and comparison groups” (Stuart, 2010, p. 11). Three balance measures were used to check covariate balance (Rubin, 2001):

- The difference in the means of the propensity scores in the two groups being compared must be small. Rubin (2001) suggests that the standardized differences of means should be less than 0.25.
- The ratio of the variances of the propensity score in the two groups must be close to 1.0. Rubin (2001) suggests that the variance ratios should be between 0.5 and 2.0.
- The ratio of the variances of the residuals of the covariates after adjusting for the propensity score must be close to 1.0. Rubin (2001) suggests that the variance ratios should be between 0.5 and 2.0.

Table C1 and C2 show the results of match diagnosis before and after the matching. Balance check confirmed that the matched comparisons were similar to the Globaloria participants in all aspects of selected covariates.

Table C1. Diagnosing Matches: Before the Matching

Variable	Participant		Comparison		Balance Diagnosis				
	<i>n</i> = 769		<i>n</i> = 14,677		Mean Difference	Pooled Variance	Standardized Mean Difference	<i>p</i> Values	Variance Ratio
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
Propensity Scores	0.07	0.06	0.05	0.03	0.02	0.03	0.67	<0.001	1.90
Pretest Covariates									
2010 Math	656.30	45.52	646.05	59.15	10.26	58.54	0.18	<0.001	0.77
2010 Reading	483.04	43.82	475.16	58.88	7.88	58.22	0.14	<0.001	0.74
2010 Science	605.68	47.07	599.60	59.40	6.08	58.84	0.10	0.005	0.79
2010 Social Studies	409.76	32.55	403.30	38.48	6.46	38.20	0.17	<0.001	0.85
Demographic Covariates					Chi-square test			<i>p</i> value	Ratio
High School	78.6%		86.1%		$\chi^2 = 32.14$ (1)			<0.001	0.91
Sixth Graders	1.2%		1.1%		$\chi^2 = 76.74$ (6)			<0.001	1.09
Seventh Graders	5.8%		10.0%						0.58
Eighth Graders	6.8%		10.3%						0.66
Ninth Graders	3.6%		6.8%						0.53
Tenth Graders	29.4%		20.0%						1.47
Eleventh Graders	26.3%		22.7%						1.16
Twelfth Graders	26.8%		29.1%						0.92
Male	65.0%		51.3%		$\chi^2 = 54.90$ (1)			<0.001	1.27
Low-Income	43.3%		39.3%		$\chi^2 = 4.86$ (1)			0.03	1.10
LEP	4.0%		6.0%		$\chi^2 = 0.67$ (1)			0.41	0.67
Disability	5.2%		10.7%		$\chi^2 = 23.73$ (1)			<0.001	0.49
Black Students	3.4%		5.6%		$\chi^2 = 6.70$ (1)			0.01	0.61
Asian Students	1.0%		1.0%		$\chi^2 = 0.05$ (1)			0.83	1.00

Table C2. Diagnosing Matches: After the Matching

Variable	Participant		Comparison		Balance Diagnosis				
	n= 701		n = 701		Mean Difference	Pooled Variance	Standardized Mean Difference	<i>p</i> Values	Variance Ratio
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>					
Propensity Score	0.06	0.03	0.06	0.03	0.00	0.03	-0.01	1.00	1.00
Pretest Covariates									
2010 Math	654.74	45.22	654.66	49.47	0.08	47.36	0.00	0.98	0.91
2010 Reading	481.60	43.86	481.30	47.25	0.30	45.55	0.01	0.90	0.93
2010 Science	604.76	47.09	606.38	50.70	-1.62	48.89	-0.03	0.54	0.93
2010 Social Studies	408.11	32.40	407.43	35.39	0.68	33.90	0.02	0.71	0.92
Demographic Covariates					Chi-square test			<i>p</i> value	Ratio
High School	82.3%		83.3%		$\chi^2 = 0.25 (1)$			0.62	0.99
Sixth Graders	1.0%		1.0%		$\chi^2 = 32.93 (6)$			<.001	
Seventh Graders	8.3%		7.0%						1.19
Eighth Graders	8.4%		8.7%						0.97
Ninth Graders	7.1%		2.6%						2.73
Tenth Graders	21.1%		29.7%						0.71
Eleventh Graders	24.0%		27.4%						0.88
Twelfth Graders	30.1%		23.7%						1.27
Male	63.8%		67.3%		$\chi^2 = 1.97 (1)$			0.16	0.95
Low-Income	42.7%		45.5%		$\chi^2 = 1.16 (1)$			0.28	0.94
LEP	0.1%		0.0%		$\chi^2 = 1.00 (1)$			0.32	1.00
Disability	5.7%		4.7%		$\chi^2 = 0.71 (1)$			0.40	1.21
Black Students	3.7%		2.4%		$\chi^2 = 1.94 (1)$			0.16	1.54
Asian Students	0.7%		0.6%		$\chi^2 = 0.11 (1)$			0.74	1.17

Appendix D

Hierarchical Linear Modeling: Impact Model

Appendix D: Hierarchical Linear Modeling: Impact Model

This appendix presents the HLM models and technical report for each student WESTEST2 outcome. Table D1 shows the definition of the parameters for level 1 and 2 models.

Table D1. Level-1 and Level-2 Models: Notations and Definitions

Level 1 Equation	
$y_{ij} = \beta_{0j} + \beta_{1j} * (\text{Participation}_{ij}) + \beta_{2j} * (\text{Grade}_{ij}) + \beta_{3j} * (\text{Gender}_{ij}) + \beta_{4j} * (\text{LowIncome}_{ij}) + \beta_{5j} * (\text{Participation} \times \text{Grade}_{ij}) + \beta_{6j} * (\text{Participation}_{ij} \times \text{Gender}_{ij}) + \beta_{7j} * (\text{Participation}_{ij} \times \text{LowIncome}_{ij}) + \beta_{8j} * (\text{Math@2010}_{ij}) + \beta_{9j} * (\text{Science@2010}_{ij}) + \beta_{10j} * (\text{Reading@2010}_{ij}) + \beta_{11j} * (\text{SocialStudies@2010}_{ij}) + v_{ij}$	
Notations	
y_{ij}	Dependent variable (student WESTEST2 score) for case i in school j
β_{0j}	The intercept (student mean WESTEST2) in school j
$\beta_{1j} - \beta_{11j}$	Regression coefficient (slope) in school j
v_{ij}	Level 1 residual for case i in school j
Level 1 predictors	
$\text{Participation}_{ij}$	Globaloria participants = 1 Comparisons = 0
Grade_{ij}	Students from 9th to 12th grade = 1 Students from 6th to 8th grade = 0
Gender_{ij}	Male = 1 Female = 0
LowIncome_{ij}	Students receiving free or reduced-price meal status = 1 Students without free or reduced-price meal status = 0
$\text{Participation}_{ij} \times \text{Grade}_{ij}$	Interaction term of Participation and Grade
$\text{Participation}_{ij} \times \text{Gender}_{ij}$	Interaction term of Participation and Gender
$\text{Participation}_{ij} \times \text{LowIncome}_{ij}$	Interaction term of Participation and Income
Math@2010_{ij}	2010 WESTEST2 math (Grand mean centered)
Science@2010_{ij}	2010 WESTEST2 science (Grand mean centered)
Reading@2010_{ij}	2010 WESTEST2 reading (Grand mean centered)
$\text{SocialStudies@2010}_{ij}$	2010 WESTEST2 social studies (Grand mean centered)
Level 2 Equations	
$\beta_{0j} = \gamma_{00} + \gamma_{01} * (\text{HighPovertySch}_j) + \gamma_{02} * (\%2010\text{MathProficiency}_j) + \gamma_{03} * (\%2010\text{ReadProficiency}_j) + u_{0j}$	
$\beta_{1j} = \gamma_{10} + \gamma_{11} * (\text{HighPovertySch}_j) + \gamma_{12} * (\%2010\text{MathProficiency}_j) + \gamma_{13} * (\%2010\text{ReadProficiency}_j)$	
$\beta_{2j} = \gamma_{20}$	
$\beta_{3j} = \gamma_{30}$	
$\beta_{4j} = \gamma_{40} + \gamma_{41} * (\text{HighPovertySch}_j) + \gamma_{42} * (\%2010\text{MathProficiency}_j) + \gamma_{43} * (\%2010\text{ReadProficiency}_j)$	
$\beta_{5j} = \gamma_{50}$	
$\beta_{6j} = \gamma_{60}$	
$\beta_{7j} = \gamma_{70}$	
$\beta_{8j} = \gamma_{80}$	
$\beta_{9j} = \gamma_{90}$	
$\beta_{10j} = \gamma_{100}$	
$\beta_{11j} = \gamma_{110}$	
Notations	
γ_{00}	The population intercept (student mean WESTEST2 score)
$\gamma_{10} - \gamma_{110}$	The population regression coefficient (mean slope) for the regression of the dependent variable on the level 1 predictor
$\gamma_{01} - \gamma_{03}$	The population regression coefficient for the regression of the dependent variable on the level 2 predictor
$\gamma_{11} - \gamma_{13}$	The population regression coefficient for the interaction between the level 1

$V_{41} - V_{43}$	and level 2 variables in predicting the dependent variable (i.e., cross-level interaction terms).
u_{0j}	Level 2 residual for school j
Level 2 Predictors	
HighPovertySch _{j}	When a school had more than 50% of students receiving free or reduced-price meal status = 1 When a school had less than 49% of students receiving free or reduced-price meal status = 0
%2010MathProficiency _{j}	Percentage of students scoring proficient or higher in math in school j
%ReadProficiency _{j}	Percentage of students scoring proficient or higher in reading in school j

Note. Notation follows Raudenbush and Bryk (2002).

Tables D2-D5 show the results of the two-level HLM model examining the effect of Globaloria on student WESTEST2 achievement outcomes. Model 3 is the final model for each WESTEST2 outcome.

Table D2. HLM Results for Two-Level Model Examining the Effect of Globaloria on Student WESTEST2 Math Outcome

Parameter Estimates	Null Model		Model 1		Model 2		Model 3	
	β	SE	β	SE	β	SE	β	SE
Mean of Math Intercept (β_0)	658.00***	2.68	638.38***	11.57	636.85***	12.13	623.93***	9.15
High Poverty School			-0.78	4.53	-0.57	4.31	2.89	5.74
% 2010 Math Proficiency			0.51*	0.25	0.42	0.25	0.61	0.32
% 2010 Reading Proficiency			-0.15	0.34	-0.10	0.32	-0.04	0.35
Mean of Participation Slope (β_1)					5.00	5.94	36.37**	12.26
High Poverty School							-7.66	6.06
% 2010 Math Proficiency							-0.32	0.36
% 2010 Reading Proficiency							-0.26	0.39
Mean of Grade Slope (β_2)			8.59*	3.50	12.17*	4.79	12.94*	5.47
Mean of Gender Slope (β_3)			4.86*	1.91	5.42*	2.54	5.24*	2.51
Mean of Low-Income Slope (β_4)			-32.64*	14.66	-30.65	15.81	-29.42	15.16
High Poverty School			11.25*	5.24	11.67*	5.08	11.72*	5.20
% 2010 Math Proficiency			-0.23	0.36	-0.24	0.34	-0.24	0.35
% 2010 Reading Proficiency			0.70	0.42	0.69	0.41	0.70	0.40
Mean of Math @ 2010 Slope (β_8)			0.37***	0.04	0.37***	0.04	0.37***	0.04
Mean of Reading @ 2010 Slope (β_9)			0.13**	0.04	0.13**	0.04	0.13**	0.04
Mean of Science @ 2010 Slope (β_{10})			0.11**	0.03	0.11**	0.03	0.11**	0.03
Mean of Social Studies @ 2010 Slope (β_{11})			0.18**	0.06	0.19**	0.06	0.18**	0.06
Participation X Grade (β_5)					-6.94	5.61	-11.17	6.43
Participation X Gender (β_6)					-1.53	3.98	-0.64	3.95
Participation X Low-Income (β_7)					-4.43	4.66	-5.06	4.77

Parameter Estimates	Null Model		Model 1		Model 2		Model 3	
	β	SE	β	SE	β	SE	β	SE
Variance Components								
Between school variation (var. of u_{0j}, τ_{00})	282.45***		12.860		0.33		0.93	
Within school variation (var. of γ_{ij}, σ_{00})	1870.41		1161.98		1168.18		1161.68	
No. of parameters	3		16		20		23	
Deviance (FIML)	10324.53		9794.73		9790.60		785.52	
Intra-Class Correlation Coefficient (ICC)	13.1%		1.1%		0.03%		0.08%	
Within school variation explained (level-1 R^2)			37.9%		37.5%		37.9%	
Between school variation explained (level-2 R^2)			95.4%		99.9%		99.7%	

Note. Robust estimations are reported. Missing values are handled by using FIML. The calculation of R^2 is based on Raudenbush & Bryk's version (2002).

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

Table D3. HLM Results for Two-Level Model Examining the Effect of Globaloria on Student WESTEST2 Science Outcome

Parameter Estimates	Null Model		Model 1		Model 2		Model 3	
	β	SE	β	SE	β	SE	β	SE
Mean of Science Intercept (β_0)	618.12***	3.99	609.96***	14.62	610.75***	15.73	611.69***	16.82
High Poverty School			-9.64	6.30	-0.49	6.41	-11.01	7.80
% 2010 Math Proficiency			-0.04	0.36	-0.07	0.37	1.38**	0.47
% 2010 Reading Proficiency			-0.02	0.42		0.42	-1.24*	0.49
Mean of Participation Slope (β_1)					-4.54	5.17	5.52	20.33
High Poverty School							1.44	10.95
% 2010 Math Proficiency							-2.35***	0.57
% 2010 Reading Proficiency							1.87**	0.6
Mean of Grade Slope (β_2)			28.02***	5.57	26.42***	7.06	22.34**	7.38
Mean of Gender Slope (β_3)			0.12	3.60	-2.22	4.73	-2.09	4.74
Mean of Low-Income Slope (β_4)			-8.98	19.99	-5.25	21.43	-5.64	21.38
High Poverty School			14.59**	4.70	14.65**	4.75	14.54**	5.06
% 2010 Math Proficiency			-0.51	0.33	-0.55	0.33	-0.53	0.35
% 2010 Reading Proficiency			0.44	0.49	0.44	0.49	0.44	0.49
Mean of Math @ 2010 Slope (β_8)			0.12**	0.04	0.12**	0.04	0.11**	0.04
Mean of Reading @ 2010 Slope (β_9)			0.21***	0.06	0.21***	0.06	0.21***	0.06
Mean of Science @ 2010 Slope (β_{10})			0.27***	0.05	0.26***	0.05	0.27***	0.05
Mean of Social Studies @ 2010 Slope (β_{11})			0.24**	0.08	0.25**	0.08	0.25**	0.08
Participation X Grade (β_5)					3.85	8.15	5.43	10.14
Participation X Gender (β_6)					5.12	6.60	4.65	6.51
Participation X Low-Income (β_7)					-4.82	4.95	-4.77	5.24

Parameter Estimates	Null Model		Model 1		Model 2		Model 3	
	β	SE	β	SE	β	SE	β	SE
Variance Components								
Between school variation (var. of u_{0j}, τ_{00})	733.7***		126.34***		129.84***		84.69***	
Within school variation (var. of γ_{ij}, σ_{00})	1652.6		1124.99		1119.88		1117.44	
No. of parameters	3		16		20		23	
Deviance (FIML)	6593.5		6297.08		6295.13		6282.22	
Intra-Class Correlation Coefficient (ICC)	30.75%		10.10%		10.40%		7.00%	
Within school variation explained (level-1 R^2)			31.90%		32.20%		32.70%	
Between school variation explained (level-2 R^2)			82.80%		82.30%		85.10%	

Note. Robust estimations are reported. Missing values are handled by using FIML. The calculation of R^2 is based on Raudenbush & Bryk's version (2002).

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$

Table D4. HLM Results for Two-Level Model Examining the Effect of Globaloria on Student WESTEST2 Reading Outcome

Parameter Estimates	Null Model		Model 1		Model 2		Model 3	
	β	SE	β	SE	β	SE	β	SE
Mean of Reading Intercept (β_0)	493.98***	2.79	489.75***	11.20	490.65***	8.95	476.95***	9.77
High Poverty School			-2.71	4.48	-2.90	3.81	-3.03	4.95
% 2010 Math Proficiency			-0.71*	0.30	-0.68**	0.20	-0.20	0.27
% 2010 Reading Proficiency			1.02**	0.34	0.98***	0.24	0.87**	0.23
Mean of Participation Slope (β_1)					-0.29	5.26	36.25**	13.57
High Poverty School							-1.51	7.00
% 2010 Math Proficiency							0.09*	0.39
% 2010 Reading Proficiency							0.05	0.39
Mean of Grade Slope (β_2)			-3.57	3.89	-2.79	4.73	-6.26	5.66
Mean of Gender Slope (β_3)			-13.59***	2.41	-17.20***	3.03	-17.07***	3.02
Mean of Low-Income Slope (β_4)			-4.30	13.64	-9.79	11.30	-9.32	10.18
High Poverty School			1.54	5.14	2.85	4.43	3.85	4.23
% 2010 Math Proficiency			0.03	0.38	0.00	0.38	0.00	0.39
% 2010 Reading Proficiency			-0.05	0.45	0.02	0.45	0.02	0.45
Mean of Math @ 2010 Slope (β_8)			0.1**	0.03	0.1**	0.04	0.10*	0.04
Mean of Reading @ 2010 Slope (β_9)			0.55	0.03	0.55***	0.09	0.54***	0.08
Mean of Science @ 2010 Slope (β_{10})			0.00	0.03	0.00	0.04	0.001	0.04
Mean of Social Studies @ 2010 Slope (β_{11})			0.39**	0.05	0.385***	0.08	0.38***	0.08
Participation X Grade (β_5)					-1.99	5.45	-1.37	7.31
Participation X Gender (β_6)					7.68	4.58	7.69	4.59
Participation X Low-Income (β_7)					7.31	4.37	5.99	4.25

Parameter Estimates	Null Model		Model 1		Model 2		Model 3	
	β	SE	β	SE	β	SE	β	SE
Variance Components								
Between school variation (var. of u_{0j}, τ_{00})	278.55***		38.88**		27.29*		8.97	
Within school variation (var. of γ_{ij}, σ_{00})	2347.74		1103.84		1099.31		1098.99	
No. of parameters	3		13		20		23	
Deviance (FIML)	10127.30		9377.08		9367.08		9355.05	
Intra-Class Correlation Coefficient (ICC)	10.6%		3.4%		2.4%		0.8%	
Within school variation explained (level-1 R^2)			53.0%		53.2%		53.2%	
Between school variation explained (level-2 R^2)			86.1%		90.2%		96.8%	

Note. Robust Estimations are reported. Missing Values are handled by using FIML. The calculation of R^2 is based on Raudenbush & Bryk's version (2002).

*** $p < .001$

** $p < .01$

* $p < .05$

Table D5. HLM Results for Two-Level Model Examining the Effect of Globaloria on Student WESTEST2 Social Studies Outcome

Parameter Estimates	Null Model		Model 1		Model2		Model 3	
	β	SE	β	SE	β	SE	β	SE
Mean of Social Study Intercept (β_0)	411.43***	1.52	401.87***	5.97	405.01***	7.12	400.21***	9.26
High Poverty School			3.99	2.62	3.69	2.72	2.90	3.81
% 2010 Math Proficiency			-0.18	0.16	-0.22	0.16	-0.03	0.19
% 2010 Reading Proficiency			0.36	0.18	0.37*	0.19	0.33	0.28
Mean of Participation Slope (β_1)					-3.64	4.11	9.43	9.11
High Poverty School							1.25	3.61
% 2010 Math Proficiency							-0.36	0.21
% 2010 Reading Proficiency							0.02	0.28
Mean of Grade Slope (β_2)			-4.65*	2.28	-5.02	3.29	-6.77*	3.21
Mean of Gender Slope (β_3)			3.64	2.44	2.41	2.95	2.30	2.97
Mean of Low-Income Slope (β_4)			5.40	9.02	3.38	9.10	3.90	9.04
High Poverty School			-4.43	3.50	-4.19	3.61	-3.90	3.64
% 2010 Math Proficiency			0.12	0.20	0.17	0.22	0.18	0.21
% 2010 Reading Proficiency			-0.22	0.25	-0.25	0.26	-0.24	0.26
Mean of Math @ 2010 Slope (β_8)			0.12***	0.03	0.12***	0.03	0.12***	0.03
Mean of Reading @ 2010 Slope (β_9)			0.21***	0.04	0.21***	0.04	0.21***	0.04
Mean of Science @ 2010 Slope (β_{10})			0.04	0.03	0.04	0.03	0.05	0.03
Mean of Social Studies @ 2010 Slope (β_{11})			0.34***	0.04	0.34***	0.04	0.34***	0.04
Participation X Grade (β_5)					0.05	3.41	1.06	3.57
Participation X Gender (β_6)					2.43	4.23	2.98	4.32
Participation X Low-Income (β_7)					2.00	3.18	1.11	3.24

Parameter Estimates	Null Model		Model 1		Model2		Model 3	
	β	SE	β	SE	β	SE	β	SE
Variance Components								
Between school variation (var. of u_{0j}, τ_{00})	62.41***		0.44		0.37		0.20	
Within school variation (var. of y_{ij}, σ_{00})	987.12		518.29		517.56		514.42	
No. of parameters	3		16		20		23	
Deviance (FIML)	9350.79		8696.49		8695.03		8688.89	
Intra-Class Correlation Coefficient (ICC)	6.0%		0.1%		0.1%		0.0%	
Within school variation explained (level-1 R^2)			47.5%		47.6%		47.9%	
Between school variation explained (level-2 R^2)			99.3%		99.4%		99.7%	

Note. Robust estimations are reported. Missing values are handled by using FIML. The calculation of R^2 is based on Raudenbush & Bryk's version (2002).

*** $p < 0.001$

** $p < 0.01$

* $p < 0.05$