

Wisconsin Safe and Supportive Schools (S3): External Evaluation Report Summary

Daphne Kuo, Ph.D.
and
D. Paul Moberg, Ph.D.

Population Health Institute
School of Medicine and Public Health
University of Wisconsin Madison

Report to the Wisconsin Department of Public Instruction
Under a Safe and Supportive Schools (S3) Grant #84.184 from the US Dept. of Education

November 2016



University of Wisconsin
Population Health Institute
SCHOOL OF MEDICINE AND PUBLIC HEALTH

Table of Content

	Page
Executive Summary	3
Section A. Introduction and Background	5
Section B. Evaluation of Wisconsin S3	6
Section C: Methods	8
Section D. Findings	9
1. Student Level Characteristics and Outcomes	10
2. School Characteristics and School-level Outcomes	15
3. Intervention Strategies and Outcomes	22
4. Relationship of Student Behavior and Perceptions to Suspension Rates	26
5. Limitations	31
Section D. Summary and Conclusions	32
References	35
Appendix A: Seven factors of Student Behavior and Perceptions	36

Executive Summary

This external evaluation addresses the implementation and effectiveness of strategies to reduce high suspension rates and to improve school climate in over 50 high schools that participated in Wisconsin's Safe and Supportive Schools (S3) program. Under US Department of Education funding, the Safe and Supportive Schools program had the goal of reducing suspension rates while improving conditions for learning by increasing school safety, school climate and student engagement. Schools were selected on the basis of high rates and numbers of suspension and expulsions over a three year period. The primary grantee was the State of Wisconsin Department of Public Instruction (DPI); DPI provided funding, professional development and technical assistance to high schools to develop, implement and evaluate interventions at the district and building levels. Evidence-based interventions including modified administrative policies and student oriented programs were implemented. The most common interventions were Positive Behavior Intervention Supports (PBIS), bullying prevention, Restorative Justice, school climate initiatives and pupil non-discrimination policies.

The S3 program and its evaluation used **annual on-line cross-sectional surveys** of 9th and 11th grade students to assess perceptions of school safety, climate, commitment to school, and negative behaviors (violence, bullying, and AODA use). We also accessed **school-level data** on suspensions, testing results, graduation rates, and program implementation. Data on school-level suspensions, graduation rates, and standardized test scores came from DPI reporting systems. The design (Figure 1) was multi-level, with 4 waves of sequential cross-sectional student data (annual $n \sim 22,000$) nested within schools ($n \sim 55$). Comparison high schools ($n=64$) were selected using propensity score analysis and were examined for school-level outcomes, with adjustments for characteristics of the schools (such as percent non-white, percent low income, percent English as a second language, school size) to increase comparability.

In the **student self-reported (OYRBS) data**, using the student as the unit of analysis, there was significant change overall in the desired direction (i.e., a reduction of negative behavior or perceptions) on experience of violence, alcohol and drug use, bullying and harassment, and general perception of safety (Table 2). Changes on the dimensions of commitment to academics and perception of school discipline moved in an undesirable direction, while overall perception of school climate and support did not change significantly. The student-level changes in risk behaviors and perceptions that were significant were small (effect sizes ranging from .02 to .08 standard deviation units). Individual items shared in common in 2011 and 2013 with the overall random sample WI YRBSS show equivalent change or stability in S3 schools and statewide, with the exception of four items related to safety and violence, on which the S3 schools showed less change and higher levels of concern (Table 4).

We found **significant ($p < .01$) reduction in suspension rates** within schools (Table 5; Figure 4). The number of students suspended during the school year was reduced from an average of 25.4% of enrolled students to 11.2% of enrolled students over the four years studied (using school as the unit of analysis). This reduction was considerably more than that in the comparison high schools or statewide (Table 6). In addition, the duplicated rate of suspensions in S3 schools (total number of suspensions during the school year divided by total enrollment) averaged 74.1% in the 2009-10 baseline year, and was reduced to 30.3% in the 2012-13 school year. The numbers suggest that the typical student suspended was suspended about three times in the initial year, and only twice in the 2012-13 school year. Due to changes in reporting and in DPI websites, this number is not available for the final S3 year of 2013-14, nor have we been able to calculate it for comparison schools.

Multi-variate multi-level analysis (Table 10) suggests that *individual level* student perceptions of school climate and safety are influenced by student gender (females more positive), race/ethnicity (Hispanic students and non-Black minorities less positive than Black and White students), and grade level (9th grade more positive than 11th). On the school level, student perceptions become more negative as school size rises. Schools with higher percentages of Black and Hispanic students have on average more positive student perceptions and behaviors than schools with higher proportions of non-Hispanic White students. This is countered by the negative effect of high percentage of students in poverty/disadvantage, and school size. Language test scores (academic achievement) on the school level are predictive of more positive student perceptions. Finally, suspension rates within the schools do not reach significance in predicting overall student perceptions of safety and support as measured by this single latent variable. The lack of a significant association of suspension rates in predicting the overall (latent) student measure, once demographic and achievement variables are considered, further suggests the relative independence of suspension rates from student perceptions and experiences.

The relationship over time between student level variables and school-level administrative actions (i.e., suspension rates) was also modeled. These analyses (see Table 11) show that there is a modest **relationship between student perceptions and behaviors and suspension rates**. Schools with a higher percent of Black students have significantly higher suspension rates. Experience (and fear) of violence by students is also significantly associated with suspension rates in our models, and is a stable predictor over the four years examined. These data suggest that aside from serious violent behavior and AOD use, suspension rates are not well predicted or explained by student perceptions and behavior, but are related independently to student racial and demographic composition, even after adjusting for other variables.

Significant reductions in suspension rates reflect changes in administrative behaviors of school officials and district policy. **Reduction in suspensions has been one of the primary goals of the S3 initiative, and the data shows clear evidence of success in this regard.** Student perceptions and behaviors appear to be more resistant to change than administrative behaviors. Of ongoing concern and deserving of further investigation and analysis is the independent association of Black student race with higher levels of suspension. This is particularly important given that, controlling for other variables, our models suggest that Black and Hispanic students have overall more positive perceptions and behaviors than non-Hispanic White students in similar circumstances, but higher suspension rates.

Thus our evaluation concludes that **there was a high degree of success in reducing out of school suspension rates** in the S3 schools, and some small but **significant effects on student behavior and perceptions in the areas of bullying, AOD use, general sense of safety, and violence**. The role of student race as an independent factor related to higher suspension rates is deserving of further attention, as is further evaluation of the relative effectiveness of the various interventions which were implemented under the S3 program.

A. Introduction and Background

The Safe and Supportive Schools (S3) Program was funded by the U.S. Department of Education's Office of Safe and Drug-Free Schools in 2010. The stated goal of the program was to "...create and support safe and drug-free learning environments and to increase academic success for students in these high-risk schools." A total of \$38.8 million per year in Safe and Supportive Schools (S3) grants were provided to 11 State educational agencies (SEAs) over a four year period. According to the Dept. of Education:

Safe and Supportive Schools awards grants to State Educational Agencies (SEAs) to support statewide measurement of, and targeted programmatic interventions to improve, conditions for learning in order to help schools improve safety and reduce substance use.

(<http://www2.ed.gov/programs/safesupportiveschools/index.html>)

Funded states were expected to select schools to:

- measure school safety at the building level
- help intervene in those schools with the greatest safety needs, and
- make information on school safety publicly available.

(<http://www.ed.gov/news/press-releases/us-department-education-awards-388-million-safe-and-supportive-school-grants>)

In 2010, the Wisconsin Department of Public Instruction successfully applied for the Safe and Supportive Schools federal discretionary grant program.¹ The **goal** of Wisconsin's program was to "explore strategies in select Wisconsin high schools in order to improve school safety, enhance student engagement, and create positive school climates." Wisconsin was awarded \$3.5 million per year to help high schools. Schools were selected based on high rates and numbers of suspension and expulsions. The schools with the highest numbers of suspensions and expulsions tended to be in larger urban school districts with disproportionately high proportions of minority students and students from low income families.

The major **objectives** of the Wisconsin grant were to improve conditions for learning in high schools with high suspension rates/numbers, with a focus on:

- school safety,
- school environment,
- school engagement, and
- increased capacity for making data driven decisions.

Nineteen school districts participated, with a combined total of 52 high schools. Each district received support from the Wisconsin Department of Public Instruction (DPI) in the form of funding, professional development, and technical assistance. Schools were encouraged to use evidence-based strategies in addressing their goals.

To monitor progress of the grant's goals, a (federally required) Index of Student Behavior and School Environment (ISBSE) score was developed for each school. The Wisconsin ISBSE was compiled using a weighted combination of results from a student survey called the Online Youth Risk Behavior

¹ A 2015 Wisconsin Department of Public Instruction report entitled *Wisconsin Success Stories: Safe and Supportive Schools (S3) Grant* (Bulletin # 15047) provides a summary regarding implementation and outcomes of this grant. Much of this section is adapted with permission from that report with only minor editing.¹

Survey (OYRBS) and a school-level behavior score. Participating schools were required to survey their 9th and 11th grade students each year using this on-line survey system. The behavior score was the rate per 1,000 students of out-of-school suspensions and expulsions in each school.

S3 Interventions

Zero-tolerance policies of mandatory suspension or expulsion became generalized over the past 25 years to a wide variety of behavioral issues in schools, with the unintended consequences of a high rate of out of school suspensions in many schools. A major policy emphasis in the S3 schools was to foster a significant shift in the way schools respond to students' problematic behaviors. Schools were encouraged to distinguish between severe and non-severe behaviors, and to develop alternative strategies to suspension and expulsion, particularly for non-severe but disruptive behavior. The expectation is that such policies not only clarify discipline but help reduce the number of incidents for both severe and non-severe levels of behavior.

Participating S3 schools were also expected to utilize innovative evidenced-based interventions which had a track record of success. Assistance was provided by DPI to help schools use data on local needs to select the most appropriate interventions. Annual grantee workshops were held. Many commonly used strategies were adopted by S3 schools, including:

- Positive Behavioral Interventions and Supports (PBIS),
- Restorative Practices,
- Freshman transition programs such as Link Crew,
- SBIRT (Screening, Brief Intervention and Referral to Treatment),
- Gay Straight Alliances (GSA's),
- Classroom management programming, and
- LifeSkills Training

Many schools also developed local adaptation of strategies from these specific program models. Additionally, the critical factors of community engagement, collaboration, and sustainability were addressed through planning and technical assistance from DPI S3 consultants and other providers (such as the PBIS Network). Culturally competent programming was also a central theme.

B. Evaluation of Wisconsin S3

1. Internal Evaluation

During the first year of S3 implementation, DPI initiated an internal evaluation of the program, including development of a statewide data collection plan for the initiative. This included compiling data from internal DPI systems regarding rates of suspension and expulsion in Wisconsin high schools, and selecting the schools with the highest rates for participation. The S3 version of the DPI's Online Youth Risk Behavior Survey (OYRBS) was developed and participating schools conducted the first wave survey of their 9th and 11th grade students in 2011. DPI analysts subsequently calculated the federally required baseline "School Safety Index," which Wisconsin named the Index of Student Behavior and School Environment (ISBSE). This survey was repeated annually throughout the life of the grant. An innovative feature of the OYRBS system is the ability of local school districts to work online with their own district's OYRBS data, and DPI provided annual workshops in which one of the features was to assist schools and districts in working with their own student data. The federal goal of facilitating data-based decision making was thus explicitly addressed.

Over the four years of the grant, DPI continued internal evaluation by assessing trends in S3 schools' suspension and expulsion rates, annual OYRBS administration and analysis of change on individual items and on the ISBSE overall. In addition, success stories were collected and documented by the S3 consultants at DPI. These efforts resulted in the report recently issued by DPI which shows impressive overall change on the ISBSE index and many specific accomplishments in individual schools (DPI, 2015).

2. External Evaluation

Wisconsin DPI initially contracted with the University of Wisconsin Population Health Institute (UW-PHI) to assess the psychometric properties of the ISBSE. We analyzed the S3 student OYRBS data set to examine the internal consistency of the nine items used in the scale. We also conducted additional principal components factor analysis of the entire data set, and examined the relationship between the OYRBS items and the behavioral indicators of suspensions and expulsions (Moberg and Kuo, 2012).

The conclusions from the initial analysis were that the **survey items** in the S3 OYRBS reflect a wide range of content relevant to the goals of safe and supportive schools. Item wording is generally good, and nearly all items are derived from standardized and nationally tested survey tools from the CDC. We recommended that the survey stand as is and continue to be used.

The **“School Safety Scores”** (ISBSE) using the original algorithm developed a priori was psychometrically problematic. While the 9 survey items selected represent a range of areas across the domains of the survey, these items do not hold together psychometrically as a single measure. In fact, two of them were isolates that did not load with any other items. The reliability of the 9 items when scored as a single scale was weak ($\alpha = 0.58$). The empirical structure of the data set suggests that there are seven distinct dimensions tapped by OYRBS survey, the distribution of which may be relevant to school programming and policy decisions. By using only 9 items which did not “hang together” well, the original “School Safety Score” risked losing important information which could be beneficial to full understanding and policy development. We named the 9 dimensions General School Climate and Support; Experience of Violence; Bullying and Harassment; Alcohol and Drug Use; School Commitment/Importance of Academics; General Perception of Safety; and Perceived Rule Enforcement. Our final analysis conducted both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to refine the dimensions. These dimensions had stable reliability across groups when the sample was partitioned by sex, grade level, race/ethnicity, and academic performance of students. (Details available in our comprehensive technical report; Kuo & Moberg, 2016.)

The combination of school suspension and expulsion rates with the 9 survey items into a single “Safety Score” was also questionable. The algorithm weighting the survey items and the suspension rates gives a disproportionately heavy weight to suspensions/expulsions in the scoring. The data we presented show a huge variation in school-level rates of suspension which were not well-explained by variation in student behaviors. Our preliminary analysis yielded only modest correlations between student OYRBS responses and suspension rates. Student self-reported behaviors and perceptions of climate explained only between 12% and 35% of the variation between schools in suspension rates. It is likely that suspension rates reflected school-level administrative practice, internal policies and precedents which are conceptually very different from the constructs captured by the school's student survey responses. Given that the goal of the program is to improve school safety and supportiveness, the issue of suspension and expulsion rate needed to be addressed and measured separately from the dimensions of safety and support. We recommended that these areas be decoupled for reporting and

program development and evaluation purposes, although as a grant requirement the ISBSE would continue to be used.

We thus **recommended that the original S3 OYRBS survey stand as is and continue to be used.** While continuing to use the existing survey, including its use in calculation of the School Safety Score, we also recommended that the additional scaling based on the derived factors would be a useful component for purposes of program planning and evaluation over the course of the grant program.

Subsequently, The DPI contracted further with the UWPHI to continue to work with the S3 data, including providing annual reports to the participating schools, to conduct additional rigorous analysis of the large data set which was generated during the initiative, and to prepare this final external evaluation report. The present document summarizes the results of our work, which is presented in more detail in a companion technical report (Kuo and Moberg, 2016).

Our analyses had two major goals: documenting and reporting progress in school safety and environment performance among participating schools; and comparing the effectiveness (outcomes) of the S3 program and strategies relative to changes in non-participating schools. In addition to statistical tests for the significance of differences in changes over time, we employed advanced statistical techniques such as propensity score matching, multivariate regression and multilevel analyses to provide a more scientifically rigorous evaluation.

C. Methods

The overall design was to compare key outcomes over time and between the S3 schools and a comparison group of schools in the state. The student level outcomes were measured by annual S3 OYRBS surveys from 2011 to 2014. The school level outcomes were derived from annual data routinely reported by all Wisconsin public schools to the Department of Public Instruction. Comparison schools to evaluate school level outcomes (suspension rates, graduation rates, test scores, etc.) were selected using a propensity score matching approach to assure as much similarity as possible to the S3 schools². We also examined YRBS trends in Wisconsin overall using the items from the CDC's YRBS core which were included in the S3 survey.

Figure 1 shows the evaluation design, where Tx indicates S3 intervention/treatment while Oi's indicate measurement of outcomes, including both school and individual student outcomes: O^{sch}_i 's refers to school level outcomes and O^{stu}_i 's indicate student level outcomes. Only school level outcomes are available in comparison (Propensity Score Matched Comparison-PSMC) schools, since they did not implement the S3 OYRBS. Note that while Tx's were not included for the comparison group of schools, that did not mean that intervention strategies of evaluation interest had not be implemented in those schools—many schools did implement elements of PBIS without S3 funding.

² There was an additional set of several schools included in the S3 project that were originally designated by DPI as comparison schools. However, data collection from these schools was not uniform and the selection was not random, so to avoid biasing the results we have decided not to report the data from these schools in this summary report.

Figure 1: Evaluation Design

	<u>←2010</u>		<u>2011</u>		<u>2012</u>		<u>2013</u>		<u>2014→</u>
<u>S3 schools (N=55)</u>	O^{sch}_0	Tx	O^{sch}_1	Tx	O^{sch}_2	Tx	O^{sch}_3	Tx	O^{sch}_4
		Tx	O^{stu}_1	Tx	O^{stu}_2	Tx	O^{stu}_3	Tx	O^{stu}_4
<u>PSMC schools(N=64)</u>	O^{sch}_0		O^{sch}_1		O^{sch}_2		O^{sch}_3		O^{sch}_4

Analyses presented in this summary report include statistical assessment of change over time on student reported scales in S3 schools, and comparison of S3 student OYRBS responses to statewide YRBS items. On the school level, we present data on change over time in S3 schools and change in S3 schools relative to change in the propensity-score selected comparison schools. We also provide analysis of data on the relationship between student self-reports and school suspension rates (through 2014). Full methodological details are available in our comprehensive technical report (Kuo and Moberg, 2016).

D. FINDINGS

The internal report prepared by DPI (DPI, 2015) on the results of S3 in specific schools and districts, highlights successes in many areas of program implementation and outcome, in particular reductions in rates of suspension in S3 schools. After four years, the average Index of Student Behavior and School Environment (ISBSE) score dropped (improved) 41 percent in S3 schools. All schools in the S3 grant project reported policy changes in the area of dealing with student behavior. Many of the districts reported that the S3 endeavor resulted in a significant shift in the way they responded to students’ problematic behaviors, as well as a reduction in behavioral incidents.

“At the end of the 2012-13 school year, S3 project schools in Wisconsin reported a total of 541 fewer acts of endangering behaviors. Endangering behaviors are violent or threatening acts without physical injury. This is an 11 percent overall reduction, or approximately 10 fewer such acts per school per year.” (DPI, 2015)

The results that follow represent an in-depth external statistical analysis of the data generated under the S3 program. These findings are consistent with the DPI’s report, while offering more complex statistical analysis and further examination of relationships between the various outcomes and background variables.

1. Student Level Characteristics and Outcomes

In total, we have usable data from 87,120 surveys of students in grades 9 and 11; surveys were conducted annually over 4 years in S3 schools (N=53)³. In the primary analytical sample to assess change (eliminating schools not answering either the 2011 or 2014 survey), we have 19,369 students in 2011 and 22,365 in 2014 from S3 schools.

The social demographic composition of responding S3 students in OYRBS were fairly similar across all four survey years (Table 1). About 43% were White, 21% were Black, another 21% were of Hispanic origin, and about 14% were other or mixed races. There were more 9th graders than 11th graders; about 54% of students were 9th graders. The gender composition was quite similar across years, with about 51% female. The self-reported grades mostly earned were also similar across all four years for S3 students, with about 26% reporting mostly A's, 33% mostly B's, 24% mostly C's, and 11% mostly D's or F's. About 30% of responding S3 students over the four years were from the Milwaukee Public School District, and 69% were from urban districts. Of all OYRBS students, 53% were from "mega" size schools (defined by DPI as > 1500 students) and 37% were from "large" (801 to 1500 students) high schools.

We analyzed the OYRBS survey data provided by the students in the participating S3 schools over four years of survey administration. The empirical structure of the data set suggests that there are seven distinct dimensions tapped by the OYRBS survey, the distribution of which may be relevant to school programming and policy decisions. We named these dimensions General School Climate and Support; Experience of Violence; Bullying and Harassment; Alcohol and Drug Use; School Commitment/Importance of Academics; General Perception of Safety; and Perceived Rule Enforcement (see Appendix A for details on which items load in each scale). Our final analysis conducted both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) to refine the dimensions we extracted during the first year of analysis. These dimensions had stable reliability across groups when the sample was partitioned by sex, grade level, race/ethnicity, and academic performance of students.

While the structure of the data was relatively stable across demographic groups, disparities between groups were also reflected in these data. In particular, relative to White students, **Black students** experienced higher levels of violence, felt less safe, reported higher levels of alcohol and drug involvement, and perceived less consistent discipline.⁴ On a positive note, Black students reported more commitment to school work, and a more positive sense of school climate and support than White students. **Hispanic students** also reported higher levels of violence than White students, reported more bullying and harassment than either Black or White students, reported higher alcohol and drug use than either White or Black students, and did not differ in their perceptions of safety and of school climate and support.

³ Due to changes in the structure of several schools over the course of the S3 initiative, varying compliance with the annual survey requirements, and delay in availability of the 2013-14 school level data, the numbers of schools and students vary by analysis presented. We have attempted to clarify the sample n's in each table, and present more detailed information in the technical report.

⁴ Note that these relationships are bi-variate and do not control for background variables, which were included in our more complex modeling. Thus the direction of several of these disparate relationships was actually reversed when control variables were also included in the models.

Table 1: Student characteristics in S3 OYRBS Surveys, 2011-2014

Variables		2011 (N=19369)		2012 (N=21831)		2013 (N=23555)		2014 (N=22365)		All (N=87120)	
		%	SE	%	SE	%	SE	%	SE	%	SE
Race/ Ethnicity	White	45.9%	0.4%	43.0%	0.3%	42.0%	0.3%	41.9%	0.3%	43.1%	0.2%
	Black	20.7%	0.3%	22.1%	0.3%	21.3%	0.3%	21.1%	0.3%	21.3%	0.1%
	Hispanic	19.7%	0.3%	20.2%	0.3%	22.4%	0.3%	22.6%	0.3%	21.3%	0.1%
	Other	13.7%	0.2%	14.7%	0.2%	14.3%	0.2%	14.5%	0.2%	14.3%	0.1%
Sexuality	Hetero	85.5%	0.3%	83.5%	0.3%	77.7%	0.3%	79.4%	0.3%	81.3%	0.1%
	LGB	6.4%	0.2%	6.8%	0.2%	7.5%	0.2%	7.9%	0.2%	7.2%	0.1%
	NS/DK	8.1%	0.2%	9.7%	0.2%	14.7%	0.2%	12.7%	0.2%	11.5%	0.1%
Grade	11th	46.0%	0.4%	46.0%	0.3%	47.0%	0.3%	44.5%	0.3%	45.9%	0.2%
	9th	54.0%	0.4%	54.0%	0.3%	53.0%	0.3%	55.5%	0.3%	54.1%	0.2%
Gender	Male	49.1%	0.4%	48.9%	0.3%	49.6%	0.3%	49.9%	0.3%	49.4%	0.2%
	Female	50.9%	0.4%	51.1%	0.3%	50.4%	0.3%	50.1%	0.3%	50.6%	0.2%
Academic Grades	Most A	25.7%	0.3%	26.0%	0.3%	25.4%	0.3%	26.6%	0.3%	25.9%	0.1%
	Most B	33.5%	0.3%	33.1%	0.3%	33.0%	0.3%	32.5%	0.3%	33.0%	0.2%
	Most C	23.6%	0.3%	23.6%	0.3%	24.1%	0.3%	23.5%	0.3%	23.7%	0.1%
	Most D	7.5%	0.2%	7.7%	0.2%	7.6%	0.2%	7.6%	0.2%	7.6%	0.1%
	Most F	3.3%	0.1%	3.0%	0.1%	2.9%	0.1%	2.6%	0.1%	2.9%	0.1%
	DK	0.6%	0.1%	0.6%	0.1%	0.7%	0.1%	0.6%	0.1%	0.6%	0.0%
	None Above	5.7%	0.2%	6.1%	0.2%	6.3%	0.2%	6.6%	0.2%	6.2%	0.1%
Urban	Suburban/Rural	33.8%	0.3%	29.2%	0.3%	29.2%	0.3%	31.1%	0.3%	30.7%	0.2%
	Urban	66.2%	0.3%	70.8%	0.3%	70.8%	0.3%	68.9%	0.3%	69.3%	0.2%
School Size	Small (<301)	2.8%	0.1%	2.7%	0.1%	3.2%	0.1%	2.4%	0.1%	2.8%	0.1%
	Medium (301-800)	7.4%	0.2%	7.2%	0.2%	8.4%	0.2%	7.6%	0.2%	7.7%	0.1%
	Large (801-1500)	32.9%	0.3%	37.6%	0.3%	38.5%	0.3%	36.9%	0.3%	36.6%	0.2%
	Mega (>1500)	57.0%	0.4%	52.5%	0.3%	49.9%	0.3%	53.2%	0.3%	53.0%	0.2%
MPS	Non-MPS	72.9%	0.3%	68.6%	0.3%	69.4%	0.3%	70.0%	0.3%	70.1%	0.2%
	Milwaukee PS	27.1%	0.3%	31.4%	0.3%	30.6%	0.3%	30.0%	0.3%	29.9%	0.2%

There was a consistent relationship between **higher grades** in schools and relatively more positive perceptions, behaviors and experiences (as measured by these seven dimensions derived from the OYRBS). The data yield a strong linear relationship--as students report their grades as mostly A's, B's, C's, D's and F's, they report increasingly more negatively on the OYRBS measures. Students who reported a **sexual identity** as lesbian, gay or bi-sexual also showed a much more negative profile on the S3 OYRBS survey.

Student OYRBS data were analyzed to assess for significant change over time in the S3 schools. Since we do not have the same students each year, the analysis assessed for change in perceptions and behavioral reports of sequential cross-sections of students, treating them as unrelated samples. Due to the directionality of the measures, decreasing means are desirable.

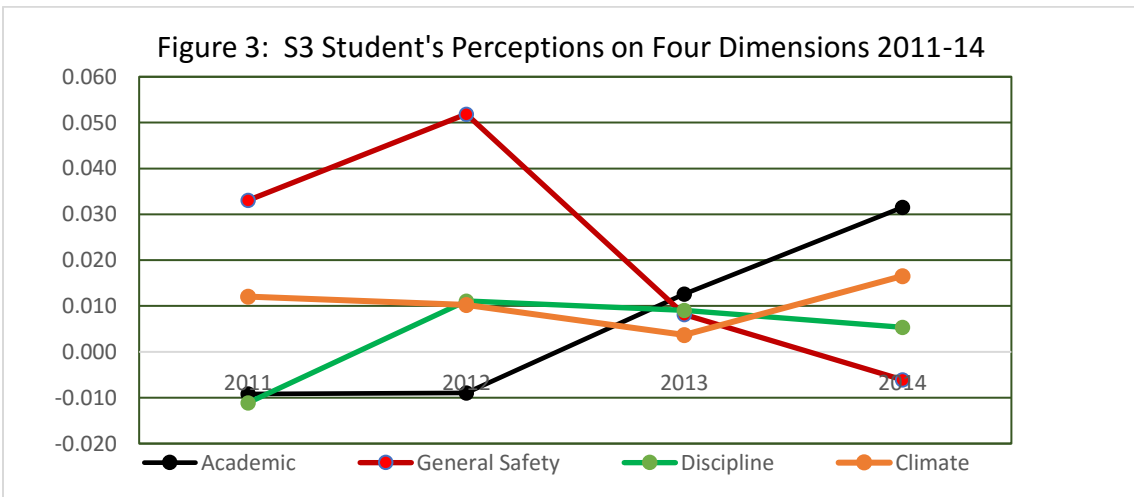
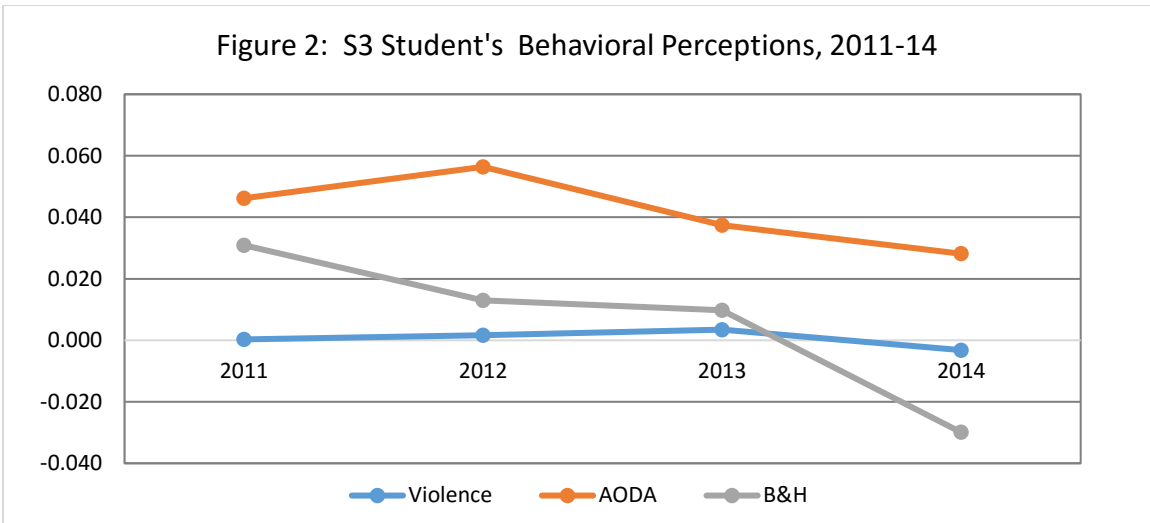
Overall, these results (see Table 2 and Figure) indicate that there were **significant changes** in the desired direction (i.e., reduction of negative behavior or perceptions) on experience of violence, alcohol and drug use, bullying and harassment, and general perception of safety. Changes on the dimensions of commitment to academics and perception of school discipline moved in an undesirable direction, while overall perception of school climate and support did not change significantly.

While statistically significant at traditional ($p < .05$) levels of significance, the effect sizes (ES = difference of means / pooled standard deviation) are very small. The effect sizes range from -.07 for Bullying and Harassment (the best result) to essentially 0 for the Supportive Climate measure (which did not register as statistically significant) to +.082 standard deviation units for Academic Commitment (the worst result). For reference, the classic text on power analysis (Cohen, 1977) suggests that effect sizes less than .20 are extremely small, potentially trivial and perhaps not of practical significance. In our case, with huge sample sizes, statistical significance is reached with very small changes. On the other hand, very small changes on the individual level may nonetheless aggregate to meaningful changes in the population (Rose, 1981, 1985).

Table 2: Student Self-Reported Behavior and Perceptions Over Time (OYRBS, S3 Students)

Year	Violence		AODA		Bullying		Academic		General Safety		Discipline		Supportive Climate	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
2011	0.000	.002	0.046	.004	0.031	.007	-0.009	.004	0.033	.005	-0.011	.004	0.012	.005
2012	0.002	.002	0.056	.004	0.013	.006	-0.009	.003	0.052	.004	0.011	.004	0.010	.005
2013	0.003	.002	0.037	.003	0.010	.006	0.013	.003	0.008	.004	0.009	.004	0.004	.005
2014	-0.003	.002	0.028	.004	-0.030	.006	0.032	.003	-0.006	.004	0.005	.004	0.017	.005
2014 vs. 2011	-0.004*	.002	-0.018*	.005	-0.061*	.009	0.041*	.005	-0.039*	.006	0.016*	.005	0.004	.007
Effect Size	-.020		-.036		-.068		.082		-.065		0.030		.006	

Note: Factors are scored with a mean of 0 for all scales. The standard deviations are 0.23, 0.52, 0.88, 0.52, 0.65, 0.53 and 0.73 respectively for violence, AODA, bullying and harassment, academic commitment, general safety, discipline and supportive climate. Scores below 0 are better/desirable, and mean scores above 0 are worse/undesirable, relative to the entire S3 population across the 4 waves of surveys. Mean change is shown in the second to last row; a negative sign is good, showing reduction in undesirable (higher score) negative perceptions, behaviors and experiences. Red/asterisk indicates significant (at $p < .05$) change/difference between 2011 and 2014. The final row estimates the effect size in standard deviation units to index the magnitude of the observed changes.



An alternative analysis of the student reported behaviors was also conducted due to the skewed nature of these variables (many students reporting few or no negative behaviors). We created a dichotomous variable reflecting student responses above the median value (across all schools and all years) for each behavioral scale. Table 3 presents the data, which again show small reductions on all scales from 2011 to 2014; the change is statistically significant ($z > 1.96$) for AODA and Bullying.

**Table 3: Percent of Students Above the 4 Year Median:
Violence, AODA and Bullying Related Behaviors and Experience**

	Violence		AODA		Bullying	
	%	SE	%	SE	%	SE
2011	49.07%	0.36%	51.06%	0.36%	51.10%	0.36%
2012	50.77%	0.34%	51.38%	0.34%	50.58%	0.34%
2013	50.45%	0.33%	50.55%	0.33%	51.14%	0.33%
2014	48.34%	0.33%	49.67%	0.33%	49.73%	0.33%
2011 vs. 2014	-0.73%	0.49%	-1.38%	0.49%	-1.37%	0.49%
z-score	1.49		2.82		2.79	

We were also able to compare change on a few OYRBS variables among S3 students from 2011 to 2013 with the statewide representative sample collected under CDC auspices for the ongoing Youth Risk Behavior Survey (YRBS). The results on variables that are common to both surveys are provided in Table 4 below. These results show the overall percentages for 2011 and 2013 for the S3 students (OYRBS) and the random sample of WI high school youth (YRBS). Comparisons are provided showing the significance of overall differences between S3 and random WI YRBS students, mean difference over time, and the difference in difference over time between S3 and YRBS samples. The latter comparison adjusted for demographic differences.

Table 4: Comparisons between Wisconsin High Schools (CDC YRBS), all S3 High Schools (OYRBS), and Milwaukee S3 High Schools (MPS)

13 questions in 2011 and 2013

Outcome Variables	Wisconsin YRBS		S3 OYRBS		S3 MPS	
	2011	2013	2011	2013	2011	2013
	N=1403	N=1452	N=19556	N=23569	N=5422	N=6767
Ever carry any weapon in last 30 days	4%	4%	5%	6%	6%	8%
Ever miss school due to feeling unsafe in the last 30 days	5%	6%	8%	10%	11%	13%
Someone ever trying to hurt you [weapon] @ school in the last 12 mos**	21%	16%	23%	23%	25%	26%
Ever threatened or injured @ school (12 months)*	6%	6%	8%	6%	8%	8%
Ever in a physical fight @ school (12 months)	12%	9%	16%	14%	23%	23%
Ever drink one drink in any day in the last 30 days	33%	31%	31%	29%	25%	25%
Ever drink 5 drinks in any day in the last 30 days	18%	16%	17%	16%	13%	13%
Ever use marijuana in the last 30 days	20%	17%	23%	23%	24%	27%
Agree or agree strongly that violence is a problem in school***	28%	21%	37%	37%	43%	41%
Not always feeling safe @school***	58%	55%	61%	66%	70%	71%
Ever offered illegal drugs @school in last 12 mo	23%	19%	23%	20%	24%	21%
Ever bullied @school in last 12 mos	24%	27%	20%	22%	13%	16%
Ever electronically bullied @school in last 12 mos	15%	19%	15%	16%	11%	10%

Difference in difference analysis of change from 2011 to 2013, adjusting for background characteristics:

*p < .10

**p < .05

***p < .01

These data show that the S3 population is not radically different from random WI students on many of these variables, although given the large sample sizes there are statistically significant differences on six of the 13 variables. Note also that Milwaukee students reported differences from both the overall S3 group and the CDC YRBS sample on several variables, in particular those related to safety and violence. Milwaukee youth also reported less alcohol use than the other groups, and less bullying. A statistical analysis called “difference in differences,” examined change from 2011 to 2013 for the overall S3 group compared to the YRBS sample, adjusted for gender, race/ethnicity, grade level and academic grades. There was a significant difference in reports of “someone trying to hurt you with a weapon at school,” with the rate declining in YRBS samples (to 16%) but staying constant (23%) in the S3 group. Being threatened or injured at school declined among S3 students to 6%, the same rate as among YRBS students. Agreement that violence was a problem at school declined (to 21%) among YRBS students but was constant at 37% among S3 students. Consistent with these other items, not always feeling safe at school decreased slightly among YRBS students while increasing slightly among S3 students. None of the other items yielded significant differences in change between YRBS and S3 groups.

Having examined the individual student level data based on student surveys over time in S3 schools, we now turn to the school level of analysis where administrative data are available on trends within S3 and other Wisconsin high schools over the four years of the S3 program.

2. School Characteristics and School-level Outcomes

Comparison between sets of schools was necessary to ensure that any changes observed in S3 schools were not an artifact of general changes in all Wisconsin schools, or in a subset of Wisconsin schools most similar to the S3 schools. Table 5 provides data (on the school level) characterizing the S3 schools relative to (a) all Wisconsin High Schools and (b) to a group of schools selected as most similar using propensity score matching (see our technical report for details).

School responses to student behavior are operationalized by annual reports on enrollment and suspension/expulsion. (Due to reporting/release delay at DPI, we obtained suspension/expulsion data for the 2013-14 school year only at the last minute as this report was being finalized, so these data are less thoroughly analyzed than suspension rates through 2012-13.) We focus on *percent of students* suspended or expelled during the year rather than total number of suspensions/expulsions, since this incidence measure was unreasonably large for some S3 schools, reflecting many students who were repeatedly suspended. In some schools, there were more total suspensions in the baseline year than there were students, indicating that some students had been repeatedly suspended.

Academic outcomes include *school-level* standardized math and reading test scores and 4 year high school completion rates. The measures of standardized tests were coded as percentage of students with **proficient or advanced** math and reading. (For reference, in 2013-14 among high school students statewide, 6% were considered “advanced,” 37% “proficient,” 40% “basic” and 17% as “minimal performance” in reading; in math the corresponding rates were 12%, 34%, 35% and 18%.

[<http://wisedash.dpi.wi.gov/Dashboard/Page/Home/Topic%20Area/Academic%20Performance/WSAS%20%28WKCE%20and%20WAA-SwD%29>]

High school 4 year completion rates, low SES (proxy= free and reduced lunch eligibility), race/ethnicity, urbanicity and school size were also assessed and used in subsequent analyses. All of these school-level variables are available publicly in the Wisconsin's Information System for Education Data Dashboard (WISEdash).

Table 5: Comparison of all High Schools in Wisconsin, S3 Schools and Propensity Score Matched Schools (School as Unit of analysis)

		Wisconsin	S3 Schools	PSMC
		N=502	N=53	N=64
Variable		%	%	%
Suspension	2010	9.1%	25.4%	8.9%
	2011	7.6%	20.9%	7.7%
	2012	7.8%	19.7%	8.4%
	2013	6.8%	14.5%	7.4%
	2014	5.7%	11.2%	7.1%
Attendance	2010	91.9%	88.0%	91.1%
	2011	92.5%	90.1%	92.0%
	2012	92.1%	88.3%	91.7%
	2013	91.6%	87.5%	90.8%
	2014	91.3%	87.7%	92.1%
Reading prof & higher	2010	35.1%	23.8%	32.9%
	2011	31.4%	20.6%	29.8%
	2012	36.0%	24.4%	33.8%
	2013	34.5%	22.1%	31.1%
	2014	36.6%	24.6%	31.9%
Math prof and higher	2010	33.4%	20.2%	31.1%
	2011	34.4%	20.5%	31.5%
	2012	34.9%	21.0%	31.9%
	2013	35.2%	21.2%	32.3%
	2014	35.6%	21.5%	31.1%
High School completion	2010	82.8%	77.4%	77.7%
	2011	84.7%	78.7%	81.6%
	2012	85.1%	78.2%	80.0%
	2013	85.3%	76.7%	79.1%
	2014	87.7%	76.7%	79.4%
English as 2nd language	2010	2.6%	7.6%	4.2%
	2011	2.7%	7.9%	4.1%
	2012	2.7%	8.1%	3.8%
	2013	2.4%	7.1%	3.3%
	2014	2.1%	6.9%	2.9%

Low SES	2010	34.3%	55.6%	40.1%
	2011	37.0%	55.0%	42.1%
	2012	38.0%	59.3%	42.2%
	2013	39.5%	60.3%	43.7%
	2014	39.7%	61.2%	45.1%
NH Black	2010	8.7%	31.7%	15.6%
	2011	8.6%	32.0%	15.4%
	2012	9.0%	32.2%	15.8%
	2013	9.2%	33.0%	15.8%
	2014	9.1%	33.0%	15.7%
Hispanic	2010	5.2%	14.6%	7.4%
	2011	5.9%	15.7%	8.2%
	2012	6.2%	16.5%	8.2%
	2013	6.8%	17.1%	8.8%
	2014	6.9%	17.7%	9.4%
Urban Schools		17.8%	64.2%	35.4%
Rural Schools		50.1%	11.3%	26.6%
Large Schools		5.7%	32.1%	8.9%
Mega Schools		20.0%	34.0%	40.5%

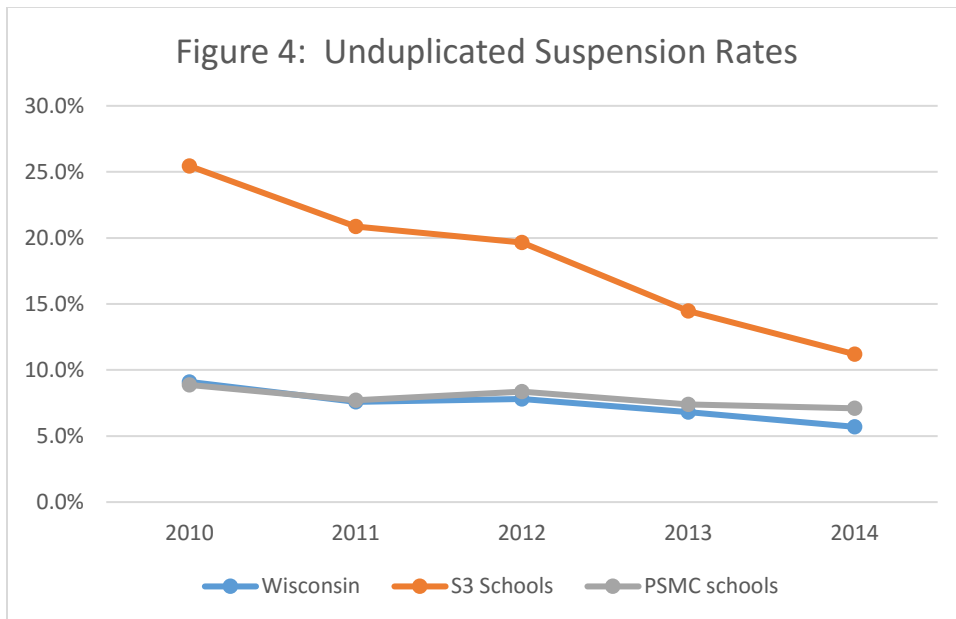
These data (Table 5) show that across all years, students in the S3 schools performed less well in both math and reading, had fewer students completing high school, and had higher suspension rates than the Wisconsin high schools in general. The S3 schools were twice or more likely to suspend/expel their students. However, the S3 schools also had a much more disadvantaged socioeconomic profile than the average Wisconsin high school. The data confirm that **selection of schools for S3 participation was appropriate to the intent of the program**; most of the S3 high schools face far more significant challenges than the typical Wisconsin high school.

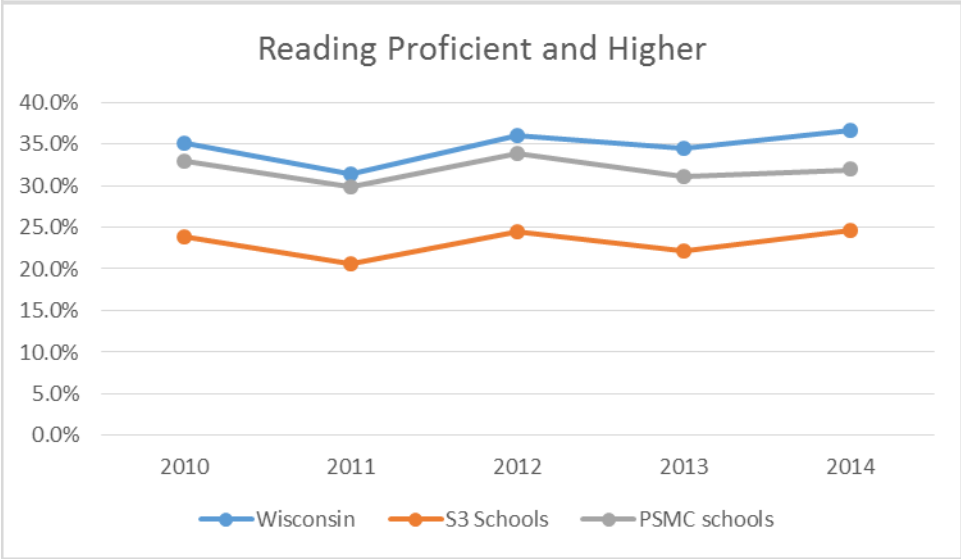
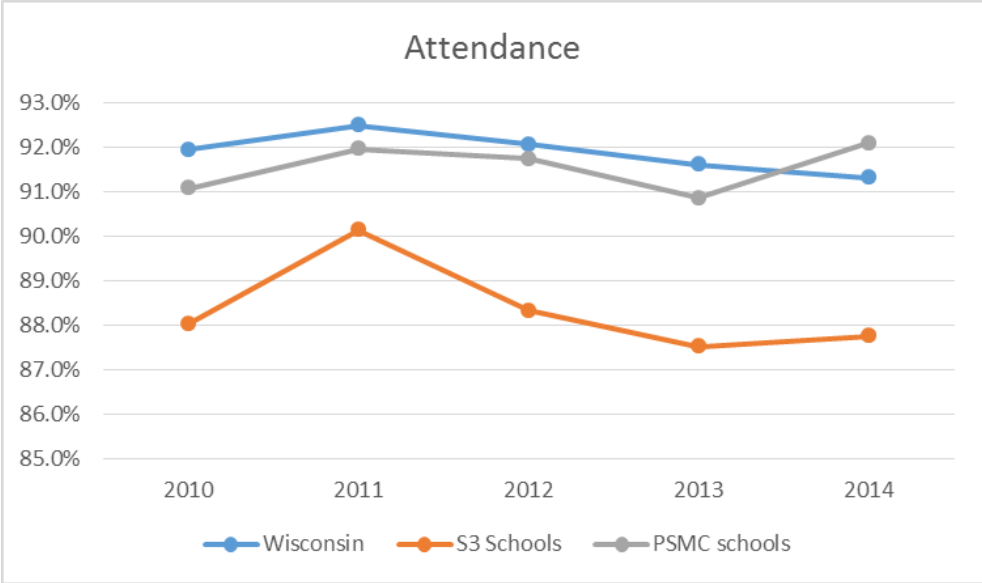
The discrepancy between S3 schools and the average WI schools further illustrate why our evaluation plan included selecting a comparison group of schools with characteristics more similar to S3 schools. We used a statistical technique called propensity score matching to select from all high schools in the state the closest matches for S3 schools. Since the majority of the larger urban high schools in the state were included in the S3 grant, finding close matches using propensity score matching (PSMC schools) was still less than ideal, although the size of differences was attenuated with the PSMC schools when compared to all high schools. Subsequent analyses, using a difference in difference regression model comparing change in suspension rates (and other school level variables) from 2011 to 2014 in S3 relative to PSMC schools, also adjusted (controlled) for baseline differences. In this way, we selected the most similar group of schools possible for comparison, and further balanced them with the S3 schools using statistical adjustment.

The graphs on the following pages illustrate the relative trends between all Wisconsin High Schools, the S3 schools, and the Propensity Score Matched Comparison (PSMC) Schools. The regression table (Table 6) summarizes our tests for significant treatment effects when change from 2010 to 2014 (2013 for suspension data) in S3 schools is compared to change in the PSMC Schools.

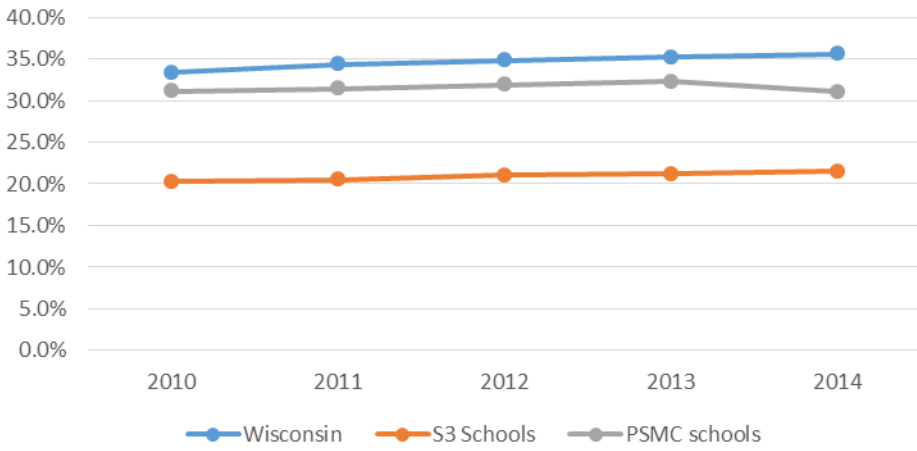
We found significant ($p < .01$) **reduction in suspension rates** within S3 schools. The number of students suspended during the school year was reduced from an average of 25.4% of enrolled students to 11.2% of enrolled students over the four years studied (using school as the unit of analysis). This reduction was considerably more than that in the comparison high schools or statewide. In addition, the duplicated rate of suspensions in S3 schools (total number of suspensions during the school year divided by total enrollment) averaged 74.1% in the 2009-10 baseline year, and was reduced to 30.3% in the 2012-13 school year (not shown in tables). These numbers suggest that the typical student suspended was suspended about three times in the initial year, and only twice in the 2012-13 school year. Due to changes in reporting and in DPI websites, this number is not available for the final S3 year of 2013-14, nor have we been able to calculate it for comparison schools. The regression analysis (Table 6) confirms that the reduction in suspension rates was significantly greater in S3 schools than it was in PSMC Schools.

Attendance rates declined slightly in S3 schools, significantly different from PSMC schools where attendance rates improved slightly. There was no difference in change in 4 year high school graduation rates between S3 and PSM schools. Reading proficiency increased slightly in S3 schools, significantly different than in PSMC schools where there was a slight decrease; math proficiency showed no effect of the S3 interventions.





Math Proficient and Higher



High School Completion

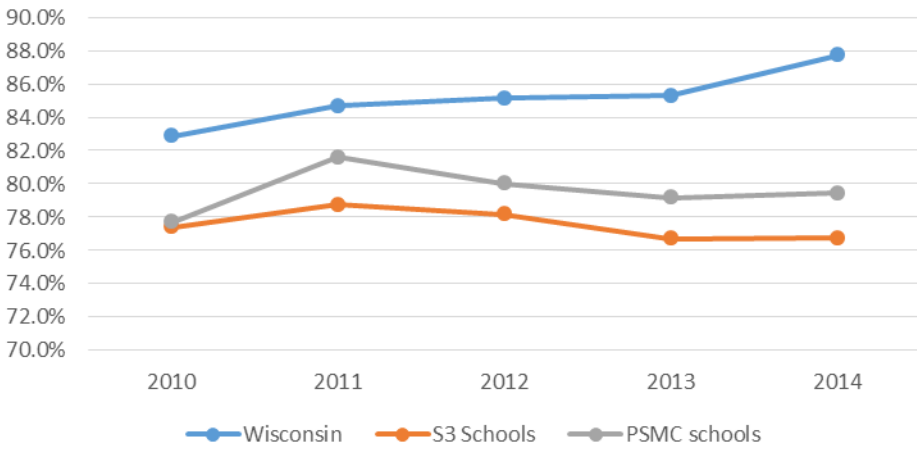


Table 6: School Level Outcomes—Difference-in-Differences Regression

Variables	School**	2010	2014	Diff	DID Regression*			
		%	(2013) %	(Change) %	Coef.	SE	z	P>z
Suspension***	S3	25.4%	14.5%	-11.0%				
	PSM	8.9%	7.4%	-1.5%	-4.7%	1.7%	-2.740	0.006
Attendance	S3	88.0%	87.7%	-0.3%				
	PSM	91.1%	92.1%	1.0%	1.9%	0.6%	3.290	0.001
High School graduation	S3	77.4%	76.7%	-0.7%				
	PSM	77.7%	79.4%	1.8%	3.5%	4.4%	0.810	0.419
Reading Prof/Advance	S3	23.8%	24.6%	0.8%				
	PSM	32.9%	31.9%	-1.0%	5.7%	2.2%	2.610	0.009
Math Prof/Advance	S3	20.2%	21.5%	1.3%				
	PSM	31.1%	31.1%	0.0%	7.8%	4.4%	1.780	0.074

* In the difference-in- difference regressions, a set of school characteristics are controlled, including proportion Black, Hispanic, other minority; male; low SES; and percent English as the second language students in each schools.

** There are 53 S3 schools and 64 PSM schools. For each S3 school, the two non-S3 schools with the closest propensity scores were included in the PSM comparison group; since the same PSM schools matched several S3 schools, the ratio is not 2:1.

***At the time of this analysis, 2013-2014 suspension data were not available. Our DID comparison thus was limited to years 2009-10 versus 2012-2013. See Table 5 and Figure 4 for the suspension rates for 2013-2014; the final rates were 11.2% (S3) and 7.1% (PSM), continuing the downward progress.

3. Intervention Strategies and Outcomes

Data regarding S3 implementation strategies were compiled from school reports to DPI, S3 consultant records based on their work in support of schools, and assessments conducted under auspices of the separate Wisconsin PBIS Network (<http://www.wisconsinpbisnetwork.org>). All schools participating in S3 were required to report on their implementation strategies to the DPI. In addition, PBIS assessment data from these schools were made available by the PBIS Network office to the evaluation team.

The evaluators worked with DPI to develop an analytic scheme to operationalize the adopted strategies among S3 schools. The strategies were coded by 1) whether they were “evidence based,” or formal “structured programs” though lacking necessary data to be considered evidence based, or other programs (e.g., locally developed approaches); 2) the focus areas of the strategies such as violence, bullying and harassment, alcohol and drugs, social emotional growth and mental health; and 3) the approach of the strategies such as individual or environmental/systems. The evidence based programs were programs with empirical data demonstrating proven effective. The structured programs were documented programs with training and technical supports but with little or no empirical evidence of effectiveness. The other programs usually were local program approaches supported by business or local agencies (such as police or local coalitions) but with no external structure (such as a manual, training, or technical support). We calculated the total numbers of strategies reported, total numbers of evidence based/individual/environmental strategies, and coded for the focal areas (such as violence, mental health AODA) within each school.

The PBIS assessment tools were inconsistently administered across schools. To create summary indices of PBIS strength, we summarized the data to reflect: 1) the average number of annual assessments (i.e., number of assessment divided by number of years in PBIS) and 2) ever passed a specified PBIS fidelity level.

Table 7 summarizes these data. Across all S3 schools over the four years, a total of 885 programs were identified and coded. For the average school, about 18 separate programs or strategies were reported or documented over the four years of S3; the large standard deviation indicates that this is highly variable, with schools reporting a very wide range of in terms of number of interventions. Most programs were structured (not evidence based), a combination of environmental and individual approaches, and with multiple purposes (focus areas). Social emotional programs were also popular. Schools participating in the PBIS movement took about 6.5 tests on average over their tenure in the program. The years in PBIS ranged from 4 to 1. The PBIS schools took about 2.2 assessments per year. About 2.5 assessments or 71% of tests taken by each school were passed (i.e., indicated fidelity with PBIS standards).

These data have significant limitations. PBIS programs were reported as one overall model for some schools, while others reported each sub-program adopted under the PBIS process separately. Milwaukee reported data for all participating schools district wide, further complicating the analysis and requiring duplication of the district reported data for each school. Thus caution should be used in interpretation of this section of the report, and the results considered as exploratory at best.

Table 7: Descriptive Statistics for Interventions including efforts in PBIS		
Variable	%/Mean within schools	SD
Self- and Consultant Reported Programs		
% of programs, evidence based	8%	
% of programs, structured	43%	
% of programs, environmental	10%	
% of programs, individual	17%	
% of programs, both env & ind	23%	
% of programs, violence	0%	
% of programs, bullying & h.	2%	
% of programs, AODA	5%	
% of programs, mental health	0%	
% of programs, SocEmo	19%	
% of programs, academic	3%	
% of programs, multiple purposes	20%	
# of evidence based programs	1.50	2.05
# of environmental programs	1.70	1.96
# of Individual programs	2.79	2.51
# of Violence programs	0.02	0.13
# of Bully/Harass. programs	0.36	0.75
# of AODA programs	0.91	1.56
# of Mental Health programs	0.02	0.13
# of SocEmo programs	2.77	1.81
# of academic programs	0.59	1.14
# of multiple purposes	3.18	2.28
Total number of programs (mean)	17.71	7.34
PBIS		
Total number of assessments	6.59	4.32
Average assessments per year	2.24	1.53
# of PBIS assessments passed	2.48	2.45
% of passed PBIS assessments	71%	46%

Analysis of the effectiveness of various approaches to intervention is difficult since there is no fixed pattern or set of interventions that were implemented consistently and with fidelity in multiple schools. Further, schools selected interventions that met their own unique set of circumstances, without a systematic experimental design. Thus our exploratory analysis is limited to assessing whether any of the individual intervention characteristics appears to have influenced average outcomes (as measured on the school level). The outcomes examined include the full set we have considered throughout this evaluation, including both academic, behavioral, and student perceptual variables, although the interventions generally did not explicitly target this full set of potential outcomes. The correlational results provided in Table 8 suggest that there were some significant contributions to student and school level change in these outcome dimensions.

Table 8: Correlation of Change in Outcomes with Program and Implementation Variables^{1,2,3}

		Suspension (Student)	General Safety	School Work	Violence	AODA	Bullying/ Harassment	Climate	Discipline
Total # of Programs	γ	-0.041	0.224	0.259	0.199	-0.078	0.230	0.224	0.330
# of PBIS assessments	γ	-0.218	0.109	0.228	-0.142	-0.315	0.168	0.181	0.016
PBIS fidelity	γ	-0.161	0.083	0.082	-0.022	-0.184	0.126	0.009	0.017
Evidence Base									
# evidence based programs	γ	0.028	0.195	0.062	0.029	-0.121	0.110	0.025	0.284
# of structured programs	γ	-0.229	0.257	0.318	0.190	-0.081	0.147	0.356	0.252
Environmental and Individual Focus									
# of environment programs	γ	-0.043	0.171	-0.003	0.150	-0.049	0.026	0.086	0.184
# of individual programs	γ	-0.256	0.218	0.325	0.053	-0.204	0.017	0.309	0.295
# of combined programs	γ	-0.002	0.192	0.200	0.140	0.035	0.293	0.169	0.165
Program Aim									
# Bullying & Harrass	γ	0.039	-0.067	-0.136	-0.171	0.003	-0.190	-0.204	-0.109
# AODA	γ	0.024	0.041	0.038	-0.131	-0.084	-0.038	0.062	0.185
# Mental Health	γ	-0.006	0.245	0.173	0.015	0.003	0.104	0.220	0.103
# Social Emotional	γ	0.058	0.034	0.218	0.106	0.133	0.199	0.268	0.251
# Academic Progs	γ	-0.067	0.222	0.151	0.235	0.011	0.037	0.058	0.253
# Multiple area prog	γ	-0.296	0.353	0.204	0.221	-0.277	0.228	0.264	0.194

1. The correlation coefficients reported here are pairwise Pearson's Correlation Coefficients between the implementation variable and school-level change from 2011 to 2014 on suspension rates and OYRBS measures.
2. Red indicates p values less than 0.1 ($p < 0.1$).
3. For suspension, a negative score is desirable; for other outcomes (7 dimensions of student OYRBS survey), a positive score is desirable. The 7 dimension outcomes in this section are reverse scored (unlike other sections of this report) for ease of interpretation.

Reviewing the data (Table 8) for each of our outcomes, we find several significant correlations between characteristics of the programs and strategies implemented in the schools and (1) change in suspension rates and (2) improvement in students' perceptions, experiences and behaviors related to school safety and supports. These correlations indicate that:

1. The efforts of the schools are correlated with the improvement. For example, the schools reporting more programs tended to make greater improvements. And the schools doing more PBIS assessments on average per year also tended to make greater improvements.
2. The number of structured programs is correlated with changes in students' perceptions of safety and supports, e.g., general safety, climate, perception of consistent discipline, and importance of school work, but uncorrelated with student's individual behaviors or experiences such as violence and AODA. It is also correlated with student suspension rates. The (rare) use of evidence-based programs is correlated with increased understanding and perception of school discipline processes.
3. The number of individual level programs is correlated with improvements in suspension rates, importance of school work, climate, and perception and understanding of school discipline. The number of environmental level programs is not correlated with any measured outcomes.
4. The number of multiple-area programs is correlated with improvement in five of eight measured outcomes (exceptions are understanding of school disciplines, experiencing violence, and importance of school work). The relationship between the multiple area program with improvement in AODA experience and behaviors is negative. Single area programs did not appear to be as effective at the school level as the multiple-area program in improving the measured outcomes, although there were some isolated areas of promise, for example social emotional programs were positively related to improved climate and perceptions of discipline.
5. To improve bullying and harassment in school, a "holistic" approach seems more effective than others. For example, the number of combined-approached programs and the number of multiple area program are the only two program variables associated with the improvement.
6. Area specific programs, such as programs targeted at bullying and harassment or AODA, are not effective in improvement of their own areas in this data set.

These results must be considered cautiously since there was no control over which schools implemented which programs, there was no good check for program fidelity and we are only reporting associations without any control for characteristics of schools and students. Coding of programs was also imprecise given the data that were available. In addition the sample is small (49-53 schools) and the reported correlations are all relatively modest. Finally, longer term implementation of programs (e.g., AOD prevention has been ongoing in schools since the 1980s) was not accounted for. The relationships noted above are all deserving of further research and analysis with stronger theory and methodology.

4. Relationship of Student Behavior and Perceptions to Suspension Rates in S3 Schools

A final set of analyses tie together the results on the school level with those from the student perspective within the S3 schools. The goal is to assess how the student perceptions and school characteristics relate to the school level responses to student behavior, i.e., suspension—and the reverse. These analyses were initially conducted prior to the availability of 2013-14 suspension rates, and were originally reported in Moberg, Kuo and Fernan (2014), appended to the technical report. The final analyses summarized here used complex multi-level and multi-variate modeling to study the extent to which students’ perceptions and experiences were influenced by school-level variables as well as individual student characteristics. Subsequently we examine the relationship between suspension rates—our primary outcome of interest—and school and student variables. What factors are associated with the rate of suspensions in the S3 schools?

First, we examined the overall raw correlations between rates of suspension and student self-reported behaviors and perceptions. These data are shown in Table 9. These results indicate strong relationships each year between rates of suspension and perceptions of violence (correlations averaging about $r=.60$ over the three years, indicating that 36% of the mutual variation is explained); consistent discipline (average $r=.52$, explaining 27% of the mutual variation); alcohol and drug use (average $r=.38$) and perception of safety (average $r=.35$). Also of interest is that the scales measuring bullying and harassment, commitment to school, and general school climate (except in 2011) have no significant relationship to suspension rates. Finally, change in these relationships over time may be important. Suspension rates appear to be consistently related to level of violence reported by students, and to perceptions of consistent discipline. However, the relationship with perception of general safety is attenuated over time (as suspension rates drop), while the relationship to reported levels of alcohol and drug use increased over time. Importantly, the relationship between violence and suspensions held fairly constant over time.

Table 9: Pearson Correlations between Suspension rates and 7 Student Reported Dimensions (N= 50-54 S3 Schools)*

Scale:	Annual Suspensions:			
	2011	2012	2013	2014
F1 Violence	0.627	0.663	0.602	0.555
F2 Bullying & Harassment	-0.018	0.025	-0.006	0.207
F3 Alcohol and Drugs	0.264	0.324	0.368	0.561
F4 Commitment to School	0.229	0.087	0.162	0.055
F5 General Safety	0.382	0.398	0.289	0.317
F6 Discipline	0.592	0.521	0.554	0.398
F7 School Climate	0.268	0.174	0.195	0.060

*Coefficients in blue font, $p < .10$

In addition, we examined the relationship between demographic variables on the school level and suspension rates across the four years of study. The significant correlations with suspension rates

include Milwaukee (MPS) school district ($r=0.49$), percent male students enrolled (0.41), percent students with free lunch (0.54); percent Black (0.66); and average test scores in language, math and reading (mean $r= .62$). Many of these variables cluster within the Milwaukee school district.

To adjust for problems in the analyses we employed multilevel models to estimate the unique influence of school environment on students’ personal experiences and perceptions of school safety and supports. This approach also accounts for explanatory variables which may be correlated with one another and with the school environment. For example, individual student’s race/ethnicity was correlated with the school-level racial/ethnic composition and school-level socioeconomic status (as measured by percent eligible for free lunch). Inner city schools also have more minority students and more socioeconomic disadvantaged students. The solution when we pool individual and school-level explanatory variables together and with the measured outcomes is to estimate a **multilevel model**, with students (level 1) clustered in schools (level 2).

A two-level Multiple Indicator-Multiple-Cause (MIMIC) model was estimated in which student perceptions are combined as a multiple-indicator latent dependent (endogenous) variable. Table 10 reports the estimated coefficients and the t-values (i.e., Coef/S.E.) including the pooled model (2011-2014) and models for each year. Given our sample size, the t-values should be greater than 1.96 in order to reach a p value greater 0.05, (i.e., the traditionally accepted probability indicative of statistical significance). Significant (at $p < .05$) explanatory coefficients are highlighted in blue in the table. Individual loadings of the seven factors are omitted from this table for readability.

Table 10. MIMIC Model Estimated Coefficients and T-value* Predicting Overall Latent Variable of School Safety and Support

	2011-2014		2011		2012		2013		2014	
	Coef	Coef/S.E.	Coef	Coef/S.E.	Coef	Coef/S.E.	Coef	Coef/S.E.	Coef	Coef/S.E.
Within Level (Student Level)										
Coefficients of Student Level Characteristics										
FEMALE	-0.008	-0.829	-0.039	-2.061	-0.039	-1.807	0.015	0.955	0.027	1.305
HISPANIC	0.097	6.602	0.106	3.538	0.091	3.044	0.131	4.627	0.071	2.429
BLACK	-0.015	-0.981	0.013	0.330	-0.047	-1.406	-0.009	-0.224	0.008	0.272
OTHRACE	0.080	4.463	0.093	2.706	0.084	2.150	0.079	2.519	0.075	1.880
Grade 9	-0.095	-6.210	-0.026	-0.817	-0.099	-3.664	-0.116	-3.938	-0.132	-4.088
Between Level (School Level)										
Coefficients of School Level Characteristics										
Milwaukee	0.090	1.133	-0.012	-0.178	0.266	1.424	-0.005	-0.036	0.101	0.869
Enrollment	0.021	3.108	0.022	2.814	0.031	2.623	0.024	1.825	0.020	1.545
% Black	-0.527	-2.616	-0.412	-1.573	-0.757	-1.942	-0.617	-1.475	-0.306	-0.789
% Hispanic	-1.145	-4.525	-0.962	-1.446	-1.382	-2.270	-0.909	-1.921	-0.404	-0.896
% Other race	-1.191	-1.353	-1.811	-0.381	0.977	0.645	0.967	0.594	-0.564	-0.311
% Free lunch	0.563	2.478	0.586	2.615	0.552	1.031	0.448	1.049	-0.636	-0.904
% Prof/Adv language	-1.118	-5.003	-0.446	-1.355	-1.448	-3.830	-0.638	-2.327	-2.281	-4.090
% Student suspension	0.408	1.717	0.201	0.315	0.679	1.012	0.945	0.672	0.906	1.222

*T-value is the coefficient/s.e.; t's of 1.96 or more are considered significant; significant coefficients of substantive interest are highlighted in blue.

Table 10 provides a summary analysis of students' perceptions and experiences regarding school safety, support and student behaviors, expressed as a composite latent variable. The loadings of dimensions on the general concept of school safety and support (not shown) suggest that the relative importance of the seven dimensions of school safety and support are consistent over the program periods. The climate scale (connectedness and support) is most important to student's rating of school safety and supports on the composite or latent factor. This latent variable is jointly predicted by individual and school level variables. The results in the table suggest the following:

1. Consistently over four years of the S3 program, Hispanic and Non-Black minority students gave significantly worse ratings, and/or have worse experiences, on the safety and support scales than Black and non-Hispanic white students. The unfavorable views declined slightly over time among non-Hispanic non-white students.
2. Black students rated the overall safety and support experiences similarly to their non-Hispanic counterparts, when other personal and school-level factors are held constant in the model.
3. In the first two years of the program, female students tended to express a more favorable view of school safety and support and more positive behavioral experiences. The gender differences disappeared in the last 2 years.
4. 9th graders tend to have a more favorable view and experience of school safety and support; the difference between them and the 11th graders increased over time.

In the 2nd panel of Table 10, we report the school-level coefficients of the overall composite variable of safety and support. The loadings (not shown) suggest more than one latent factor of school safety and support for these 7 dimensions (analysis not reported here indicated that the 7 dimensions may be separately explained by two latent factors: behaviors/experiences and perceptions.) The school level results indicate that:

1. Students from Milwaukee schools did *not* perceive or experience safety and support differently from students in other schools, other variables held constant.
2. Students from larger schools reported significantly worse experiences and perception than those from smaller schools; but these differences were only during the first two years of S3 programs, perhaps indicating a positive program effect in large schools.
3. Students from schools with a higher percentage of Non-Hispanic Black or Hispanic students tended across years to report more favorable views and experiences than those from schools with fewer Black or Hispanic students.
4. Students from schools with higher rates of free lunch reported a less favorable view only seen in 2011 and in the cumulative sample. In all other years this SES proxy was not significant.
5. Students from schools with better performance on standardized language testing tend to rate school safety and support more favorably.
6. Student suspension rate was **not** significantly related to the students' perceptions and experiences of safety and support overall nor over time.

A final set of multi-variate analyses sought to explain variation in suspension rates, rather than in student perceptions, using both the background variables and the student perceptions and behaviors. This analysis used the school-level data set (n= 205 school by year units of analysis) to assess the relationship between the various explanatory variables, student reported perceptions and experiences, and our variable of most interest, suspension rates.

Four models were estimated. The final specification of each model included year as a predictor variable to assess change in suspension rates as a function of the year of survey and hence S3 program outcome. The models were (1) year and student perception/behavior variables only; (2) year and background variables only; (3) the full set of variables including a code for MPS and MMSD; and (4) the full data set excluding codes for school districts. The school district codes were included to control for the school district confounding pointed out earlier. District codes were irrelevant for the model of MPS only, and for the overall model the coding for MMSD was insignificant and was dropped. Table 11 provides these data.

Model	All Schools (N=205)			
	1	2	3	4
Independent Variables				
2012	-0.029	-0.018	-0.017	-0.017
2013	-0.094***	-0.058***	-0.071***	-0.071***
2014	-0.150***	-0.109***	-0.132***	-0.131***
MPS		-0.004	0.023	
MMSD		-0.008	0.035	
% Black		0.165**	0.112	0.153***
% Hispanic		-0.058	-0.040	-0.012
% NW/NB/NH		-0.005	-0.067	-0.033
Enrollment/100		-0.001	-0.003	-0.003
% Male		0.717***	0.648***	0.669***
% Free Lunch		0.083	0.083	0.072
% Language proficient/advanced		-0.143**	-0.134*	-0.120*
F1: Violence	1.925***		0.775*	0.767*
F2: Bullying & Harassment	-0.661***		-0.267**	-0.252*
F3: AODA	-0.281***		-0.042	-0.051
F4: Commitment	0.660***		0.590***	0.519***
F5: General Safety	0.250***		0.072	0.065
F6: Discipline	0.320**		-0.219	-0.188
F7: Climate connectedness	-0.393***		-0.079	-0.075
Constant	0.237***	-0.153	-0.080	-0.105
R-squared	0.549	0.650	0.683	0.679

*** p<0.01, ** p<0.05, * p<0.1

With $n=205$ (school by year units of analysis), we see that between 55 and 68 percent of the variation (R-squared) in suspension rates is explained by the variables included in the models. All four models yield significant negative coefficients for school years 2012 and 2013, confirming that suspension rates were reduced significantly over the course of the S3 program.

Model 1 also includes the seven student perceptual and behavioral factors. We see that all of these are related to suspension rates, when background and school-level variables are not included in the model. (Recall that these variables are scored such that negative scores are high, positive/desirable scores are low.) The largest coefficient is for the violence scale, where higher rates of perceived violence in the schools are related to higher rates of suspension. Lack of commitment to school work, unsafe perceptions on the school safety scale, and perceived lack of consistent discipline are all related to higher suspension rates. However, higher rates of bullying and harassment, higher AOD Use, and more positive climate appear to be related to lower suspension rates.

All schools Model 2, which includes only year and school background variables, shows that most of the variation (65%) in suspension rates is explained by background factors alone. Three variables reach significance—percent African American students, percent male students, and lower standardized test scores (represented by Language skills).

Model 3 includes all of the variables in one model. Here we see significant coefficients for percent male, lower language tests scores, perceptions of violence, and lack of commitment to school work. Bullying and Harassment reduces the suspension rates. However, Model 3 includes a school district variable that, while not significant, appears to have accounted for at least some of the variation. When the same model is run but without school district codes (Model 4), we see essentially the same set of significant variables except that percent black again becomes a significant predictor of suspension rates.

Importantly, given the coalescing of a number of variables within the Milwaukee (MPS) district, our final models were also run for MPS S3 schools only, and All Non-MPS S3 schools. Details are provided in our technical report; here we point out only the substantive differences found when examining Milwaukee separately from the remaining schools. For Milwaukee, in Model 1 (without controlling for background variables) violence and lack of school commitment are related to higher suspension rates among schools within MPS. Schools with better (lower) scores on climate, and with worse levels of bullying and harassment, have lower suspension rates. The only significant demographic factor (Model 2) is percent male, associated with higher suspension rates. In the saturated Model 4, percent male, low language test scores, high violence, low bullying, low commitment to school, and perception of consistent discipline are related to higher suspension rates. *Not* significant are race/ethnicity, enrollment size, free lunch, AOD use, general safety and climate.

For non-MPS S3 schools, the safe and supportive factors in model 1 are fairly weak in relationship to suspensions relative to the full set of schools or MPS alone. Violence is negatively related to suspensions—higher perceived violence is associated with lower suspension rates; higher AOD use increases suspension rates, and inconsistent discipline is associated with increased suspension rates. The background variables alone are more explanatory, with percent Black the strongest predictor of suspension rates, while higher test scores and larger size reduce suspension rates. In the saturated model (Model 4) percent Black again increases suspension rates, as does lower commitment to school work. Percent Hispanic decreases the rate; large enrollment also decreases the suspension rate, as does the level of bullying.

5. Limitations

This evaluation does have a number of limitations. This was not an experimental test of a prescribed set of interventions in the schools, and we were not able to successfully group the various interventions into discrete packages. Coding of strategies in each school was problematic, particularly for the Milwaukee District which reported all its schools together in a district wide summary. Even for programs that have a prior evidence base, the data on fidelity with which they were implemented is sparse.

While DPI initially designated several schools as comparisons, these schools differed significantly from the S3 participating schools, and several of them failed to collect all the needed data, so that comparison was inadequate. While we created a quasi-experimental comparison group of high schools using propensity scoring, this was limited in that the most comparable schools in Wisconsin were all participating in S3. The propensity-score selected comparison schools were in a middle tier of schools, different from the overall Wisconsin high school but also different from the typical S3 school. This problem was dealt with by statistically adjusting for additional contextual variables to better balance the design. We do not have data on programs implemented in these schools, but it is likely that a number of them had also implemented PBIS or other programs which were also used in the S3 schools. Hence the results favoring S3 schools are likely conservative in that some S3-like strategies were also implemented in some comparison schools.

There were also issues of several schools being restructured, closed, or combined during the course of the S3 initiative, and of missing data from both schools and students. We were not able to estimate the rate of completion of the OYRBS in each participating school at each time point; given the large Ns we are confident that the data are representative.

The unit of analysis is also problematic in evaluations such as this. On the one hand, the school is the unit of analysis for variables such as suspension, attendance and graduation rates, where small schools receive equal weight to very large schools. For the student survey measures, the sample sizes are very large and it is likely that the results are driven by students in the largest schools and districts. While multi-level modeling that uses both levels of data was incorporated in some of our analyses, this did not fully ameliorate this limitation. Further analysis could potentially partition the sample, and analyze data from the larger urban schools separately from that from the remaining high schools. As our results for Milwaukee relative to the other schools in S3 demonstrate, there are very different issues in large urban districts than in other settings; this can produce unique findings that are not generalizable statewide.

Another caution concerns the measures themselves. While we have affixed labels to each of the seven dimensions of school safety and support that were measured in the OYRBS survey, the reader should examine the actual language of the items in each scale (see Appendix) to fully understand the scope of each measure and avoid simplifying the constructs based on our labels.

Given these limitations, we have been cautious in interpretation of the results. Some results (such as the reduction in rates of suspension among S3 schools) are very pronounced and indisputable. Other findings are more tenuous, and should be considered as exploratory in lieu of stronger data which may be generated in the future.

D. Summary and Conclusions

This external evaluation addresses the implementation and effectiveness of strategies to reduce high suspension rates and to improve school climate in over 50 high schools under Wisconsin's Safe and Supportive Schools (S3) program. Under the US Department of Education funding, the Safe and Supportive Schools program had the goal of improving conditions for learning by improving school safety, school environment and school engagement. Schools were successfully selected on the basis of high rates and numbers of suspension and expulsion over a three year period. The primary grantee was the state of Wisconsin Department of Public Instruction (DPI); DPI provided funding, professional development and technical assistance to high schools to develop, implement and evaluate interventions at the district and building levels. Evidence-based and structured interventions including modified school discipline policies and student oriented programs were implemented. The most common interventions were Positive Behavioral Intervention Supports (PBIS), bullying prevention, Restorative Practice, and school climate initiatives.

The S3 program and its evaluation used annual on-line cross-sectional surveys of 9th and 11th grade students to assess perceptions of school safety, climate, commitment to school, and negative behaviors (violence, bullying, and AODA use). This report also included school-level data on suspensions, testing results, graduation rates, and program implementation. Student surveys included measures of perceived school climate, experience of violence, bullying and harassment, personal alcohol and drug use, commitment to school work, perceived safety, and perceived consistency of rule enforcement. Implementation data came from school consultants, the PBIS network and annual school reports. Data on school-level suspensions, graduation rates, and standardized test scores came from DPI reporting systems.

The design was multi-level, with 4 waves of sequential cross-sectional student data (annual n~22,000) nested within schools (n~55). Comparison high schools (n=64) were also selected using propensity score analysis and were examined for school-level outcomes, with adjustments for characteristics of the schools to increase comparability.

We found significant reduction in suspension rates within schools, with suspensions reduced from an average of 25.4% to 11.2% of students, over the four years studied. This reduction was considerably more than that in the comparison high schools or statewide. In the student self-report data, there was also significant change in the desired direction (i.e., reduction of negative behavior or perceptions) on experience of violence, alcohol and drug use, bullying and harassment, and general perception of safety. Changes on the dimensions of commitment to academics and perception of school discipline moved in an undesirable direction, while overall perception of school climate and support did not change significantly. The student-level changes in risk behaviors and perceptions that were significant were small (effect sizes ranging from .02 to .08 standard deviation units). Individual items shared in common in 2011 and 2013 with the overall random WI YRBS show equivalent change or stability in S3 schools and statewide, with the exception of four items related to safety and violence on which the students in S3 schools showed less change and higher levels of concern.

When change from 2010 to 2014 in S3 schools was compared to change in the propensity score matched comparison (PSMC) schools, the analysis confirms that the reduction in suspension rates was significantly greater in S3 schools than it was in PSMC Schools. However, attendance rates declined slightly in S3 schools, significantly different from PSMC schools where attendance rates improved

slightly. There was no significant difference in change in 4 year high school graduation rates between S3 and PSM schools. Reading proficiency increased slightly in S3 schools, significantly different than in PSMC schools where there was a slight decrease; math proficiency showed no effect.

Exploratory analyses were conducted on the school level regarding the effects of programs implemented. The total number of programs implemented was not associated with change in suspension rates but with improved perceptions of safety, commitment to school work, reduced bullying, consistent discipline and improved supportive climate. PBIS involvement was positively related to reduced suspensions, but negatively related to changes in AOD use and in student commitment to school. Use of evidence-based programs was associated with improved perceptions of discipline. Structured programs had positive results on suspensions, as well as student perceived safety, commitment to school work, discipline and climate. While environmental programs were not significantly associated with any of our outcomes, individual programs were associated with positive effects on suspension, commitment to school, discipline and climate. Programs specifically targeting behaviors had only spotty effects, although social-emotional oriented programs were associated with improved climate and perception of consistent discipline. Multi-aim programs appear to have broader effects, including reduced suspension rates and improved perceptions of safety, bullying and climate. Schools with more multi-aim programs also had less decline in AOD use. These relationships are all deserving of further analysis with stronger theory and methodology.

Analyses using multi-level modeling and a composite (latent) measure of student perceptions and behaviors, suggests that *individual level* student perceptions of school climate and safety are influenced by student gender (females more positive), race/ethnicity (Hispanic students and non-Black minorities less positive than Black and White students), and grade level (9th grade more positive than 11th). On the school level, student perceptions become more negative as school size rises. Schools with higher percentages of Black and Hispanic students have on average more positive student perceptions and behaviors than schools with higher proportions of non-Hispanic White students. This is countered by the negative effect of high percentage of students in poverty/disadvantage, and school size. Language test scores (academic achievement) on the school level are predictive of more positive student perceptions. Finally, suspension rates within the schools do not reach significance in predicting overall student perceptions of safety and support as measured by this single latent variable. The lack of a significant association of suspension rates in predicting the overall (latent) student measure, once demographic and achievement variables are considered, further suggests the relative independence of suspension rates from student perceptions and experiences.

The relationship over time between student-level variables and school-level administrative actions (e.g., suspension rates), was further modeled. These regression analyses confirm the reduction in suspension rates over time and yield a modest relationship between student perceptions and suspension rates. Overall (when a variable differentiating MPS is not included), schools with a higher percent of Black students have higher suspension rates, even after controlling for other variables. Outside of Milwaukee, schools with higher percentages of Hispanic students had a lower suspension rate. Experience and concern of violence by students is significantly associated with high suspension rates in our models, except in the model where Milwaukee is excluded. Perception of consistent discipline is also associated with suspension rates, although in one of our analyses seems to attenuate over time as suspension rates decrease, perhaps due to the more nuanced and less predictable discipline being implemented as school suspension policies changed. Alcohol and drug use rates are also related to suspension rates, with the correlational data suggesting that this relationship increased over time as suspension rates dropped. Suspension rates are thus most strongly related to student race

(Black), to student reports of experience of violent behavior, and to AOD use. These data suggest that rates of violent behavior and substance use partially predict suspension rates, and (outside of Milwaukee) suspension rates are independently related to student racial composition, even after adjusting for other variables. In Milwaukee, where the vast majority of students are individuals of color, experiences of violent behavior and lack of commitment to school are the strongest predictors of suspension. Milwaukee out of school suspension rates continue to be significantly higher than those elsewhere in the state, but have also decreased significantly.

Significant reductions in suspension rates reflect changes in administrative behaviors of school officials and district policy. This has been one of the primary goals of the S3 initiative, and the data shows clear evidence of success in this regard. Student perceptions and behaviors appear to be more resistant and slow to change than administrative behaviors. Of ongoing concern and deserving of further investigation and analysis is the independent relationship between Black race and higher rates of suspension. This is particularly important given that, controlling for other variables, our models suggest that Black and Hispanic students have overall more positive perceptions and behaviors than non-Hispanic White students in similar circumstances.

References

- Cohen, Jacob, 1977. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press.
- Kuo, Daphne and Moberg, D Paul, 2016. Technical Report: Wisconsin Safe and Supportive Schools (S3) External Evaluation. Madison WI: University of Wisconsin Population Health Institute.
- Moberg, DP, Kuo, D and Fernan, S. 2014. "Reducing High School Suspension Rates and Improving School Climate: Wisconsin's Safe and Supportive Schools Program." Presented at the annual meeting of the Society for Prevention Research, Washington DC, May 28, 2014.
- Rose, Geoffrey, 1981. "Strategy of prevention: lessons learned from cardiovascular disease." British Medical Journal 282:1847-51.
- Rose, Geoffrey, 1985. Sick individuals and sick populations. International Journal of Epidemiology 14(1): 32-38.
- Wisconsin Department of Public Instruction, 2015. Wisconsin Success Stories: Safe and Supportive Schools (S3) Grant (Bulletin # 15047).

Appendix A: Questions for Seven Factors and Measurement Structure

OYRBS Q. No.	Question
Factor 1: Violence	
879	During the past 30 days, on how many days did you not go to school because you felt you would be unsafe at school or on your way to or from school
880	During the past 30 days, on how many days did you carry a weapon such as a gun, knife, or club on school property?
881	During the past 30 days, on how many days did you carry a gun on school property?
884	During the past 12 months, how many times has someone threatened or injured you with a weapon such as a gun, knife, or club on school property?
885	During the past 12 months, how many times were you in a physical fight on school property?
886	During the past 12 months, how many times has someone tried to hurt you by hitting, punching, or kicking you while on school property?
Factor 2: Harassment and Bullying	
889	During the past 12 months, have you ever been bullied on school property?
890	During the past 12 months, have you ever been electronically bullied? (Include through e-mail, chat rooms, instant messaging, websites, or texting.)
891	During the past 12 months, have you ever bullied someone else on school property?
892	During the past 12 months, have you ever been harassed on school property?
893	During the past 12 months, have you been harassed on school property because of your race or ethnic background?
894	During the past 12 months, have you been harassed on school property because of your weight, size, or physical appearance?
896	During the past 12 months, have you been harassed on school property because someone thought you were gay, lesbian, or bisexual?
929	During the past 12 months, have you been harassed on school property because of your gender?
Factor 3: AOD Use	
899	During the past 30 days, on how many days did you have at least one drink of alcohol?
900	During the past 30 days, on how many days did you have 5 or more drinks of alcohol in a row, that is, within a couple of hours?
902	During the past 30 days, how many times did you use marijuana?
903	During the past 30 days, how many times have you taken an over-the-counter drug to get high?
904	During the past 12 months, how many times have you attended school under the influence of alcohol or other illegal drugs, such as marijuana or cocaine?
905	During the past 12 months, has anyone offered, sold, or given you an illegal drug on school property?
Factor 4: Commitment to School Work	
918	How often do you feel that the school work you are assigned is meaningful and important?
919	How interesting are most of your courses to you?
920	How important do you think the things you are learning in school are going to be for your later life?
921	Over the past school year how often did you enjoy being at school?
922	Over the past school year how often did you hate being at school?
Factor 5: General Safety	
888	Do you agree or disagree that violence is a problem at your school?
897	Do you agree or disagree that harassment and bullying by other students is a problem at your school?

Factor 6: Consistency of School Discipline

- 925 Do you agree or disagree that you understand the rules for student behavior and conduct at this school?
- 926 Do you agree or disagree that this school has clear consequences for breaking the rules?
- 927 Do you agree or disagree that staff at this school enforce the rules for student behavior and conduct?

Factor 7: Supportive School Climate

- 906 Do you agree or disagree that your teachers really care about you and give you a lot of encouragement?
- 908 Do you agree or disagree that you feel like you belong at this school?
- 924* Do you agree or disagree that this school has a friendly and welcoming atmosphere?
- 909 In my school, students have lots of chances to help decide things like class activities and rules.
- 910 There are lots of chances for students in my school to talk with a teacher one-on-one.
- 912 There are lots of chances for students in my school to get involved in sports, clubs, and other activities outside of class.
- 913 There are lots of chances to be part of class discussions or activities.
- 914 My teacher(s) notices when I am doing a good job and lets me know about it.
- 915 The school lets my parents know when I have done something well.
- 916* I feel safe at my school.
- 917 My teachers praise me when I work hard in school.

*Items also cross listed with factor 2 (harassment and bullying)