THE URGENCY OF DOING:
EVALUATING THE VALIDITY OF AN IMPLEMENTATION AND
SUSTAINABILITY MEASURE FOR SCHOOL-WIDE PREVENTION PROGRAMS

By
DOMINIC C. MOCERI

A thesis submitted to the
Graduate School-New Brunswick
Rutgers, The State University of New Jersey
in partial fulfillment of the requirements
for the degree of
Master of Science
Graduate Program in Psychology
written under the direction of
Maurice J. Elias
and approved by

________________________
________________________
________________________

New Brunswick, New Jersey
January 2011
School-based prevention and promotion interventions (SBPPI) teach students how to recognize and manage emotions, solve problems effectively, establish positive relationships with others, and develop prosocial attitudes. When implemented effectively, SBPPI have been shown to improve desirable outcomes (e.g., commitment to community, standardized achievement test scores, and attendance) and to reduce undesirable outcomes (e.g., suspensions, drug and alcohol use, and aggressive and violent behavior). Unfortunately, our understanding of how to effectively implement and sustain SBPPI outside of well-controlled conditions is lacking. In order to help build a science of implementation and sustainability, this thesis presents a conceptual framework and a measurement tool for effective SBPPI implementation. The framework differentiates among various phases of implementation, ecological levels surrounding implementation in schools, and factors in the system of implementation that facilitate fidelity and sustainability. This framework is measured by the “Schools Implementing Towards Sustainability” (SITS) scale, which is designed to be “user-friendly” in field settings by being viable and scalable. Analyses from a diverse sample of 157 schools implementing
Social-Emotional Character Development (SECD), a type of SBPPI, reveal that the SITS has good reliability, good concurrent and construct validity, and promising predictive validity. The findings of the SITS may help advance both the science of school-based interventions and the science of implementation and dissemination as a whole by demonstrating how to bridge the science/“real world” gap.
Acknowledgements

I would like to acknowledge and thank the staff of the Social-Emotional Learning Lab at Rutgers, The State University of New Jersey; the Liaisons of the Developing Safe and Civil Schools (DSACS) Project; and the students and staff at schools participating in the DSACS Project. Specifically, I would like to thank my mentor Maurice J. Elias for allowing me to embark on this project and seeing it through to the end, Jazmin Reyes for being part of the initial conversation which led to the creation of the project and providing encouragement the entire time, and Donna Synder for providing feedback on the project and making sure the DSACS Liaisons and School Coordinators completed the surveys.
# Table of Contents

Abstract ...................................................................................................................... ii

Acknowledgements .............................................................................................. iv

Table of Contents .................................................................................................. v

List of Tables ........................................................................................................ vi

List of Figures ........................................................................................................ vii

Introduction ........................................................................................................... 1

Method .................................................................................................................... 25

Results ..................................................................................................................... 33

Discussion .............................................................................................................. 44

References ............................................................................................................. 54

List of Appendixes ............................................................................................... 59
List of Tables

Table 1. Surveys Counts .................................................................................... 75

Table 2. Climate Survey Demographics ......................................................... 76

Table 3. School Demographics ................................................................. 77

Table 4. Number of Schools that Increased and Decreased their SITS scores across Time ........................................................................... 78
List of Figures

Figure 1. Effects of Implementation Assistance by Experts on Implementation & Sustainability Scores ........................................ 79
Introduction

“I have been impressed with the urgency of doing. Knowing is not enough; we must apply. Being willing is not enough; we must do.” – Leonardo da Vinci

Need for School-Based Prevention and Promotion Interventions

Provisions of the No Child Left Behind (NCLB) Act of 2002 require that schools not only improve academic achievement and close the achievement gap between high- and low-performing children but also that schools establish plans for being safe and drug-free, preventing school dropout, and providing support to delinquent youth to continue their education (Collaborative for Academic, Social, and Emotional Learning [CASEL], 2005). The NCLB Act also requires that prevention programs are grounded in scientific research and have evidence of effectiveness. While the justification for these mandates are often provided on humanistic grounds, the work of James Heckman (2007), the Nobel Prize winner in Economic Sciences in 2000, shows that investment in children makes sense even from a purely economic standpoint.

One type of school-based prevention and promotion intervention (SBPPI) is Social-Emotional Character Development (SECD), which combines Social-Emotional Learning (SEL) and Character and Moral Education. In essence, SECD is the infusion of one’s “capacity to recognize and manage emotions, solve problems effectively, and establish positive relationships with others” (Zins & Elias, 2006, p. 1) with prosocial attitudes, values and judgments. “SEL programming is based on the understanding that (1) many different kinds of problem behaviors are caused by the same or similar risk factors, and (2) the best learning emerges from supportive relationships that make learning both challenging and meaningful” (CASEL, 2005, p. 5). SECD interventions are typically
implemented in schools by teachers, administrators, and school support personnel such as school psychologists, social workers, and counselors. When implemented correctly, SECD has been shown to improve desirable outcomes (e.g., mastery of academic subject material, motivation to learn, commitment to school, commitment to community, time devoted to schoolwork, standardized achievement test scores, attendance, graduation rates, and prospects for constructive employment) and to reduce undesirable outcomes (e.g., suspensions, expulsions, grade retention, drug and alcohol use, violence and aggression) (Berkowitz & Bier, 2006; CASEL, 2005; Durlak & Weissberg, 2005; Elias et al., 1997; Greenberg et al., 2003; Hawkins, Catalano, Kosterman, Abbott, & Hill, 1999; Malecki & Elliott, 2002; Wilson, Gottfredson, & Najaka, 2001; Zins & Elias, 2006; Zins, Weissberg, Wang, & Walberg, 2004).

However, it is not enough to simply have interventions that work under experimental or test conditions. We must also have methods of transporting them into the communities that need them in such a way that the adopting schools will reap the same benefits as found in published studies. Unfortunately, despite two decades of steady growth in our understanding of what works for SBPPI in general and SECD in particular (Dalton, Elias, & Wandersman, 2006; Kam, Greenberg, & Walls, 2003), our understanding of how to implement these interventions outside of well-controlled conditions, how to sustain them over time, and how to disseminate them widely is lacking (Elias, 2007; Elias, Zins, Graczyk, & Weissberg, 2003). This process has left us with a science of “what works” but not a science of “how to.” As this lack of detailed implementation knowledge applies not only to SBPPI but to mental health services in general, many in the scientific community have stated that the time has come to build a science of implementation and
dissemination in order to bridge the science/“real world” gap (Greenberg, Domitrovich, Graczyk, & Zins, 2005; Proctor et al., 2009).

In order to build a science of implementation and dissemination, methods of studying implementation, fidelity, sustainability, dissemination, and scalability will need to be developed, empirically evaluated, and improved. Like any good scientific endeavor, this must be based upon a conceptual framework with clearly defined key terms. In order to help build a science of implementation, this thesis will review the conceptual framework of implementation as it pertains to SECD and empirically evaluate an SECD implementation scale designed to be sustainable and scalable with typical community resources.

**Review of Key Implementation Constructs**

Before reviewing the conceptual framework of implementation, several key terms must be defined. Evidence-based practice (EBP) is “the conscientious, explicit, judicious use of current best evidence in making decisions about the care of [individuals]” (Sackett, Rosenberg, Gray, Haynes, & Richardson, 1996, p. 71). Implementation is “the use of strategies to introduce or change evidence-based health interventions within specific settings” (The National Institutes of Health [NIH], 2005, “Research Objectives,” para. 11). Fidelity, which is sometimes referred to as treatment integrity or adherence, is “the degree to which a proposed intervention is delivered as it was originally intended” (Yeaton & Sechrest, 1981, p. 160). Sustainability, which is sometimes referred to as institutionalization, is defined as a program’s ability to withstand the test of time, which requires that the program is “not dependent on an influential leader or a few staff members, all of whom will eventually leave the setting” (Dalton et al., 2006, p. 374).
Dissemination is “the targeted distribution of information and intervention materials to a specific public health or clinical practice audience” (NIH, 2005, “Research Objectives,” para. 11). Scalability refers to the ability of an EBP to be done in many different sites in order to reach its audience (Elias et al., 2003). Thus, EBP is what works, implementation is the sum total of how things are done, fidelity is obeying the program designers, sustainability is the durability of a program, dissemination is the communication of an intervention, and scalability is the intervention’s ability to expand.

**Conceptualizing Implementation in Schools**

Based upon the frameworks of Greenberg et al. (2005), which conceptualized SBPPI implementation in general, and Ji et al. (2008), which conceptualized SEL implementation in particular, SECD implementation can be described in terms of three categories: phases, levels and system of implementation. These three categories interact with one another to determine the intervention that is actually implemented.

**Phases of implementation.** A phase refers to the general timeframe of where a school is in terms of implementing an intervention. While the exact number of phases and their boundaries vary from researcher to researcher, all conceptualizations include a preparation phase, an initial delivery phase, and a continuing delivery phase (Adelman & Taylor, 2002; Commins & Elias, 1991; Greenberg et al., 2005; Ji et al., 2008). The preparation phase is what occurs before an intervention begins. This includes choosing an intervention, making a commitment, training personnel, forming a core committee, and planning specifically how the intervention will be done. The initial delivery phase is the early years when an intervention occurs. This includes the personnel delivering the intervention, the core committee facilitating the delivery and fidelity of the program, the
administration actually allowing time for program planning and delivery, and someone tracking data on the program’s effectiveness. The continuing delivery phase begins when the intervention has been implemented over multiple years and when aspects of the intervention become a part of the school. This includes reflecting on the success of the implementation process, making improvements for the next round, and securing financial and structural support.

There are two main reasons for keeping phases in mind. First, they set the stage for what steps and expectations are appropriate and necessary at given time for a school. Second, they help remind the school (and intervention designers) that implementation is never truly complete and that ongoing reflection and adaptive modifications to changing circumstances are required (i.e., a school that only focuses on the first two phases will not be able to maintain the program for the long haul). Thus, the phases from an ongoing cycle where the continuing delivery phase leads back to the preparation phase.

Sustainability is not a finish line but an ever renewing starting line.

**Ecological levels surrounding implementation.** Drawing upon ecological systems theory, Greenberg et al. (2005) identify 4 levels that provide the context surrounding school-based prevention and promotion interventions (SBPPI): the classroom, the school, the district, and the community. As each level is nested within another level, the differing qualities of these 4 levels will inevitably affect what programs are chosen, the pace at which students are exposed to a program, the quality of that exposure, the sustainability of that exposure, and what adaptations are made (necessarily or unnecessarily).

The next paragraph delineates an adapted and expanded version of Greenberg et al. (2005)’s levels. The classroom, school and district levels are similar in that they each
contain environmental and structural aspects. Environmental aspects capture the attitudes, knowledge and behaviors of people (i.e., students, teachers, administrators) as well as the climate. Climate refers to the quality of interpersonal interactions, including those among and between students, staff, and parents (Mattison & Abner, 2007). The structural aspect captures the organization and resources. In short, the environment describes the people and the structure describes the institution.

The classroom level includes the implementer (e.g., behaviors and beliefs), the classroom environment (e.g., classroom climate and peer relations), and the classroom structure (e.g., classroom size). The school level includes the school environment (e.g., school climate, awareness of student needs, school goals, administrative leadership and logistical support, administrative stability, teacher mobility, student attendance, detention rates, bullying rates, violence rates) and the school structure (e.g., grade configuration, school size, student demographics, student to faculty ratios, financial resources). The district level includes the district environment (e.g., awareness of student needs, district goals, administrative leadership and logistical support, administrative stability, and communication with schools) and the district structure (e.g., financial resources, number of schools in the district). The community level includes the school-community relations (e.g., school-family relations, community support and readiness, connections to non-government agencies, and financial resources), the government-community relations (e.g., local laws, state laws, national laws), and community-community relations (e.g., how people within the community treat each other, how different neighborhoods interact with each other, violence rates, drug rates, employment opportunities).
The most important reason for keeping the levels in mind is that examining the classroom level by itself is not enough. Most SBPPI are conceptualized as individual interventions that occur in the classroom, which leads some researchers to act as if fidelity was the only key component of implementation. However, all interventions occur in a particular context (e.g., the school, the district, the community), which inevitably exercises important influences (accidentally or deliberately). The ecological systems approach makes it clear that if the disincentives in an outer ecological level for implementing the intervention are not balanced or overshadowed by the incentives at more proximal levels for implementation, then the intervention will have a difficult time succeeding (Commins & Elias, 1991). Thus, all 4 levels must be in a good-enough alignment if an intervention is going to succeed in the long-term instead of just in the short-term.

**System of implementation.** The system of implementation is the network of factors that facilitate implementation by promoting both fidelity and sustainability. The term “system of implementation” is adapted from Greenberg et al. (2005)’s term “implementation support system,” which is a school-based adaptation of Chen (1998)’s model of theory-driven evaluation. While the phrase “barriers to implementation” is frequently used in the literature, the concept of a system of implementation is more theoretically accurate and pragmatically useful. A barrier implies that once that aspect is removed, fidelity and positive intervention outcomes will occur automatically, whereas a system of implementation implies “that even the strongest, most extensively evaluated program will fail without [adequate support]” (Greenberg et al., 2005, p. 13). In other
of “bad” things, but rather by creating “good” things.

**Model of sustainable implementation.** The network of factors that affect the system of implementation has gone by many names in the research literature, but they are more similar than dissimilar. The factors have been called barriers to implementation, contextual factors, effective strategies, essential elements, implementation steps, institutionalization conditions, sustainability factors and more (Berkowitz, & Bier, 2006; CASEL Safe & Sound, 2005; Commins & Elias, 1991; Elias, 2007; Greenberg et al., 2005; Ji et al., 2008). While at first glance this approach may appear to conflate the concepts of implementation and sustainability, conceptually and pragmatically there are very few, if any, differences between “implementation factors” and “sustainability factors.”

From a conceptual viewpoint, truly effective implementation strategies must be effective sustainability strategies. If a strategy really helps an intervention be implemented with better fidelity, then that strategy will also lead to the intervention having better staying power in the school. If a strategy really helps an intervention become a part of the school in an effective way, then that strategy will also increase the probably of having greater fidelity and higher implementation rates.

Of course, it is possible to develop strategies of implementation that work against strategies of sustainability and vice versa; but that would make them ineffective strategies from any integrated implementation viewpoint. For instance, some evidence suggests that initial implementation can be fostered by making an intervention relatively easy to adopt, but this relative simplicity can work against sustainability (Commins & Elias, 1991).
Such a process works against the long-term interests of both the school and intervention developer.

More pragmatically, the findings from studies that identify “implementation factors” are highly similar to studies that identify “sustainability factors.” Greenberg et al. (2005) identify inadequate principal leadership as a barrier to implementation whereas Commins and Elias (1991) identify principal support as an institutionalization condition. CASEL (2005) identifies the lack of adequate staff development and continuing support as a barrier to implementation whereas Elias (2007) identifies ongoing professional development as a sustainability factor. Ji et al. (2008) identify the presence of a steering committee as an early implementation step whereas Commins and Elias (1991) identify the presence of a steering committee as a sustainability factor. Berkowitz and Bier (2006) identify family and/or community involvement as an effective strategy whereas and Ji et al. (2008) identify communication with stakeholders (e.g., families and communities) as an ongoing sustainability factor.

Of course, there are differences between the studies that label these particular elements as “implementation factors” and those that label them as “sustainability factors;” but these differences are misleading because no two “implementation factors” studies are identical and no two “sustainability factors” are identical. The reasons for differences between studies has less to do with the overarching label for the particular elements and more to do with the study’s methodology and the particular wording chosen to summarize the findings. Arguments against combining the “implementation factors” and “sustainability factors” studies can be used just as easily against combining two “implementation factors” studies or against combing two “sustainability factors” studies.
When examined conceptually and empirically, studies on “implementation factors” and “sustainability factors” overwhelmingly address the same mechanisms, structures, and goals and arrive at the same findings. The high overlap allows us to combine these two areas of study which are usually kept separate. This combination results in a model of sustainable implementation as its system of implementation. The model of sustainable implementation should be adopted by the field because it captures all the necessary components for solid implementation and reminds schools and researchers that implementation occurs in cycles.

**Network of factors.** As there is so much overlap in the specific factors that make up the system of implementation, only the article with the most schools and the longest duration between the initial implementation phase and the measurement of the factors will be summarized: Elias (2007) conducted 21 interviews of teachers, principals, and other school staff at 15 different sites that had previously been identified as model and flagship sites in Elias et al. (1997). Notably, these interviews were conducted 5 years after the initial program delivery phase (i.e., during the continuing delivery phase), after the initial support given to the schools were no longer in place. The 5 year timeframe is important because the literature reports that it takes 3-5 years between the time that an SECD program is started and the positive changes are seen and relatively stable (Dalton et al., 2006) and that not all schools are able to sustain SECD programs in the long run (Gager & Elias, 1997).

Using field notes as well as the transcribed tapes from site visit interviews, Elias (2007) identified seven themes of sustainability. These include: (1) there is clear commitment, participation, and reinforcement of program implementation from key
school administrators; (2) a core group of individuals, in addition to the school administrator, strongly supports the program and is very involved in it; (3) program language and skills are present in multiple aspects of daily school life; (4) the program is compatible with needs and activities in the school; (5) a plan exists and is followed for perpetuating the knowledge and skills needed to implement the program; (6) there are systematic opportunities for staff to reflect upon progress of the program; and (7) information about the program’s effectiveness is collected and reflected upon, and used in planning future directions. These are quite similar to the six ongoing sustainability factors identified in Ji et al. (2008): (1) provide ongoing professional development; (2) evaluate practices and outcomes for improvement; (3) develop infrastructure to support SEL; (4) integrate SEL framework schoolwide; (5) nurture partnerships with families and communities; and (6) communicate with stakeholders.

**Integrating Phases, Levels and the System of Implementation.**

**Phase effects on the system of implementation.** The impact of each factor of the system of implementation for a given intervention’s implementation is affected both by the current phase of the intervention and the various ecological levels described above. For instance, a serious commitment to an intervention should occur during the preparation phase; and the active participation of a core group of individuals should occur during the initial delivery phase; but, the full integration of the intervention’s language and skills throughout the school will probably not occur until the continuing delivery phase. The cyclical nature of the phases means that no factor is ever permanently finished and that one cannot expect a school to have achieved everything by the end of the first initial delivery phase.
**School level system of implementation effects on the classroom level.** As for the ecological levels, all 4 levels interact to determine the amount of fidelity and sustainability an intervention will possess. A well-designed system of implementation at the school level does not guarantee that a given teacher will implement the SBPPI within his/her classroom, but it often determines the maximum level of fidelity for its classrooms as well as the sustainability of the program within the classroom. If there are too many barriers, no teacher will be able to implement the SBPPI. And, if there is not a strong enough system, few teachers will continue to implement the SBPPI, even if they do so at the beginning.

**School environment effects on the system of implementation.** Additionally, the school environment will impact the system of implementation as a whole. For instance, a positive school climate amongst staff promotes implementation by building a sense of professional community to collaborate and meet the challenges of implementing an intervention (Greenberg et al., 2005; King & Newmann, 2000). This means that the strengthening of the staff climate may be a necessity before positive student outcomes can occur. High teacher mobility, on the other hand, may decrease the system of implementation even if there is in-house staff training because so much time is spent on training for the intervention rather than executing the intervention. Better school environments should produce better implementation in the short-run, but it is also likely that better systems of implementation can improve the school environment in the long-run.

**School structure effects on the system of implementation.** The school structure may also affect the system of implementation. Sometimes the effects are clear and the
reasons are relatively unambiguous. For instance, a high child-to-teacher ratio in schools is a risk factor for problematic social behavior in children and adolescents (Substance Abuse and Mental Health Services Administration [SAMSHA], 2002). Thus, schools with high ratios may have a harder time implementing the program and achieving improvements in student outcomes. The reasons presumably lay in how much attention the teachers and other staff are able to give to each student. School resources are another example that probably affects the system of implementation. That is, schools with fewer resources will probably not be able to implement an intervention as rapidly or thoroughly.

Other times, however, the effects of the school structure on the system of implementation are not clear and the reasons for them are quite complex. For instance, school size and grade configuration (i.e., being elementary, middle, high schools), can affect implementation in a myriad of ways. Hartmann and colleagues (2006) report that small high schools have better student outcomes than large high schools. But, they also report that school size makes no difference when the school has some choice as to who enters the school. Moreover, they report that special admission schools have better school environments and academic achievements than city-wide and neighborhood admission schools. Thus, the report of Hartmann and colleagues (2006) may provide preliminary evidence that school environment (sometimes through selection of the “best” kids) is the mediating factor for school size and student outcomes.

Grade configuration is another aspect of the school structure that may or may not affect the system of implementation (once a school chooses an intervention that is age-appropriate, of course). When rating “Select SEL” programs, CASEL (2005) identified 20 for elementary schools, 12 for middle schools and 9 for high schools. Does the greater
amount of select programs for elementary school indicate that it is easier to implement
SEL interventions in that type of setting? Does it indicate that schools, funding agencies,
or researchers are less likely to see the need for SEL interventions for high school
students? Or, does it indicate something more benign? There may simply be more
programs for elementary schools because there are more elementary schools. For
instance, in New Jersey, there are 1,625 elementary schools, 539 middle schools and 501
high schools (http://www.publicschoolreview.com/).

Supporting the premise that grade configuration may have little effect on prevention
efforts are the results of a meta-analysis by Wilson, Gottfredson and Najaka (2001).
When they looked at school-based prevention interventions for delinquency/crime, drug
use, dropout/ nonattendance, and other problem behaviors (such as aggression), they
found only one significant difference for grade configuration on student outcomes: High
schools had larger effect sizes than middle schools for preventing dropout/nonattendance.
This finding, however, may have less to do with the efficacy of programs for preventing
high school dropout/ nonattendance and more to do with society’s laws and attitudes on
these issues. That is, at a certain age (usually achieved in high school) a student can
legally drop out of school whereas society always has issues with delinquency, crime,
drug use, and aggression, regardless of the person’s age. Still, there is no guarantee that
the (non-)effects of grade configuration on student outcomes will hold true for its effects
on the system of implementation.

District and community level effects on the system of implementation. The district
and community levels may also affect the system of implementation. However, there is
much less research in these areas, so only some basic examples will be provided. A
supportive district, for instance, will enhance the system of implementation (e.g., help organize staff training, provide financial resources for the intervention) at the school level while an indifferent or adversarial district will create barriers to proper implementation (e.g., not allow certain sustainability factors to be implemented, emphasize standardize test preparation so much that there is no time left for the intervention). And, of course, the community is going to have a profound effect on the district, school and classroom. Communities full of opportunities that engage young people are going to boost the intervention options of the school whereas students coming from communities stricken with poverty, violence, and/or drugs are more likely to have negative school environments and fewer school resources for implementation efforts.

**Measuring Implementation**

While a conceptual framework of implementation is necessary, it is not sufficient for developing a science of implementation. A true science of implementation requires a way of measuring implementation. Without measurement tools, there is no way of knowing what schools are actually doing. Measuring implementation allows schools and researchers to determine what aspects of an intervention are actually causing the change in identified outcomes, what aspects are in place to help secure fidelity, and what aspects are in place to help secure sustainability. Measurement can take place at all four ecological levels during all three phases. Frequent measurements at multiple levels are critical to prevent drift. “Indeed, drift is a major problem in disseminations efforts generally, and thus the evaluation and maintenance of treatment fidelity may be a core component of ongoing efforts” (McHugh & Barlow, 2010, p. 74).
**Classroom level measurement.** Implementation in the classroom can be measured many ways. One reason that classroom-level measurement is usually emphasized is because that is the primary delivery point for most school-based interventions. Therefore, fidelity usually must be measured at this level. The most objective and comprehensive way to do so is to have trained observers (blind to experimental conditions) use a structured observation system. This provides the most accurate system for knowing how the program is being utilized. Unfortunately, this method is also the most costly. Resources (time and money) have to be spent to train the observers, to get the observers to the site, and to compensate the observers for the observations. Additionally, each observer can only observe one classroom at a time, and there is no way to know if the implementer (i.e., the teacher) utilizes the intervention in the exact same manner when the observers are not there. While classroom observations are critical for measuring the efficacy of a program, they are rarely, if ever, used in regular community settings as they are cost and time prohibitive. Thus, classroom observations alone cannot be the sole foundation of an implementation science.

Other methods to measure implementation at the classroom level also exist. For instance, teachers can provide detailed self-reports of what they implemented and/or students can provide reports on what the teacher did. Reports given by teachers and students are much more cost-effective than observations by trained outside observers. Reports can also be generated in many classes rapidly. The downside to this method, however, is that their reliability and validity are often unknown. Factors likely to influence the reliability and validity of the reports are how specific and objective the questions are and what incentives the teacher has for accurate reporting. If job
performance or salary is based upon these reports, then their validity would be highly questionable. But, if they can be executed in a constructive environment with non-punitive dedication to intervention improvement with outside consultants providing feedback, then their validity should be higher.

**School level measurement.** While fidelity is usually measured at the classroom level, most (if not all) of the factors for sustaining implementation identified above describe an ideal *school* system of implementation. This is probably because most day-to-day and month-to-month factors that affect implementation occur at the school level. Schools provide support (e.g., training and reflection time) to the implementers (e.g., the teachers). Schools are also the link between the district and the classroom and between the community and the classroom.

The most comprehensive way of measuring implementation at the school level is to conduct in-depth interviews. This is the method utilized by CASEL (1995) and Elias (2007) to identify sustainability factors. Principals, teachers, and other school personnel are all good candidates for this method. Often in-depth interviews are supplemented with a review of permanent products (e.g., school policies, training curricula, and meeting) and data systems currently in use (Horner et al., 2004). The advantage of this method is that both quantitative and qualitative data can be gathered in great detail, which makes this method ideal for developing theory and discovering what really happens in schools. The disadvantages of this method are similar to the disadvantages to the school level methods above; namely, in-depth interviews cost time and money and the incentive structure surrounding the interviews can weaken the reliability and validity of the interviews. When the review of permanent products and data systems supplement the
interviews, potential biases are lowered; but good permanent products and good data systems do not guarantee that they are followed or utilized.

Another method of measuring implementation at the school level is to administer surveys. Surveys can be given to principals, teachers, and other school personnel. Hall and Hord (2005) administered their Levels of Use scale to all school personnel before and after implementation to assess staff’s attitudes and behaviors. Ji et al. (2008) created an implementation rubric where 16 items are rated on a 4-point scale; the scale measures the network of factors that promote fidelity and sustainability. The advantage of the survey method is that quantitative data can be gathered from many schools with little cost and high reliability (as the questions/items are not open-ended like the interviews described above). The disadvantage of this method is that it is still a type of self-report, which means that it suffers from the same problems as all self-report measures do and that important information can be missed (i.e., not including an important question).

**District level measurement.** Measurement can also be done at the district level; however, this is quite rare in the research literature. Nevertheless, interviews and surveys can be conducted at the district level just as they are at the school level. The participants can be either district personnel reflecting on the support they provide to their schools or school personnel reflecting on the support they receive from their district.

**Community level measurement.** While no implementation measure is likely to be given to the community itself, critical aspects of the community should be assessed through one of the other levels. Students, teachers, school administrators, and district administrators can all provide information about what it is like living in the community. Community violence is frequently assessed, for instance, by asking students what types
of events (e.g., fights, shoot-outs) they have witnessed in their neighborhood. Additionally, objective data collected by government and non-government agencies about the community (such as crime rates, unemployment rates, and racial segregation) can be gathered. However, it must be recognized that any community level measurement is going to be a measure of potential implementation moderators rather than a measure of factors supporting the system of implementation itself.

The Present Study

The current project will examine the reliability and validity of an implementation and sustainability survey for social-emotional character development (SECD). The survey is called the Schools Implementing Towards Sustainability (SITS) Scale, where the acronym SITS is intended to emphasize the goal of sustainability (e.g., the intervention “sitting” within the school rather than disappearing when external funding or support terminates). The SITS survey was developed for the current project and is intended to be sustainable, scalable, scientifically valid, and useful in making predictions about a range of school intervention outcomes (see Appendix A for the items in the SITS survey and Appendix B for the coding system for the SITS).

SITS versus CASEL rubric. The SITS survey shares many qualities with the CASEL Practice Rubric for Schoolwide SEL Implementation studied by Ji et al. (2008). They are both designed to be used by school personnel and both describe the school level system of implementation. As both are based upon reviews of similar literature, the items of the SITS and the CASEL rubric are quite similar (see Appendix C for a comparison of the items of the SITS survey and the CASEL rubric).
However, there are important differences in terms of how they are intended to be used, how they are organized, and how the assessment is completed. The CASEL rubric is intended to be completed at least twice a year by both principals and their SEL planning teams, with each member of the SEL team completing the survey. In contrast, the SITS is intended to be completed only once a year by the school SEL coordinator/team leader, who gathers input from relevant others. This lowers the amount of time and other resources needed to complete the implementation measure, which allows for the SEL planning team to spend more time on critical tasks (e.g., strengthening the current implementation).

The two measures differ in their organization. The CASEL rubric prescribes a list of 10 steps of implementation along with 6 ongoing sustainability factors. The 10 steps of the CASEL rubric are intended to be completed in the prescribed order (although the rubric acknowledges that an individual school’s order might vary). Each stage/factor has 4 levels of quality with detailed explanations about what counts for each rating. The description of each stage/factor has its own page with about 10-15 sentences describing the possible ratings.

The SITS, on the other hand, simply presents 15 sustainability factors with no prescribed steps (although the phases were considered when developing the measure). Each factor has 4 levels of quality (i.e., false to very true) without detailed descriptions distinguishing one level from another. While at first this may appear to be a downgrade in quality, this allows the SITS to be much shorter. Being shorter is a key component to building a science of implementation as brevity increases the likelihood of a measure being used in regular community settings. The lack of a prescribed order was necessary
within the context of the current survey administration (see Setting below) and may allow for greater flexibility and generalizability.

Finally, the CASEL rubric and SITS differ in their assessment methodology. The CASEL rubric was administered to school personnel at a two-day training session for SEL during the first year of the schools’ implementation. This limits what type of responses that schools can give as they were all in the initial delivery phase (perhaps some were in the preparation phase). The SITS, however, was administered to schools via the internet at the respondent’s convenience during a longitudinal action-research project. While the SITS is not immune to the potential for biased responses, the demand characteristics for giving the “right answer” should be much less when done over the internet than when done at a training session with the program implementation team present. Moreover, the schools completing the SITS had been implementing SECD for 0-4 years, which puts most of them in the initial delivery phase, some of them in the preparation phase and some of them in the early moments of the continuing delivery phase. Additionally, whereas the CASEL rubric was only completed by school personnel, the SITS was completed by school personnel (i.e., SECD Coordinators) and by outside experts working with individual schools (i.e., SECD Liaisons) (see Method section), which allows a comparison between responder types to be possible.

**SITS description.** The SITS scores represent a process indicator of how effectively the system of implementation for SECD is being delivered. Higher scores indicate the presence of a more robust network of factors that facilitate implementation by promoting both fidelity and sustainability (see Measures section and Appendix B). The 15 factors of
implementation covered by the SITS can be mapped onto the phases of implementation and onto the ecological levels surrounding implementation (see Appendix D).

The mapping of the SITS factors to the phases exists to demonstrate that all three phases of implementation are covered by the SITS and to indicate when that factor is most likely to be implemented. For instance, a school is likely to form an SECD committee during the preparation phase but unlikely to have partnered with community agencies at this time. However, there are two caveats that must be kept in mind. First, the SITS does not prescribe a specific order for each factor to be implemented. A specific order implies that there are critical periods for certain factors to be obtained. However, there is no empirical evidence to support that position or the particular order delineated. The position of the SITS is that it is important to achieve each factor, but the implications of the order and the specific and relative importance of each has not been empirically established. A cumulative effect of achieving many implementation factors may be more important than developing each factor in a predefined way. Therefore, a school would not be designated as having entered the initial delivery phase because they achieved steps 1-3 but rather because they have achieved several factors; and a school would not be designated as having entered the continuing delivery phase because they achieved steps 1-10 but rather because they have achieved the vast majority of the factors.

Second, a school never finishes implementing a factor. Since the continuing delivery phase never ends, the factors must be maintained over time in order for the intervention to be sustainable. For instance, making a serious, integrated and long-term commitment to SECD during the preparation phase accomplishes nothing if it is abandoned in either the initial delivery phase or the continuing delivery phase. And, if a school stops tracking
quantifiable data after entering the continuing delivery phase, then the school will not know when (not if) the intervention’s effects begin to weaken, indicating the need for changes to the intervention. These caveats to the mapping of phases onto factors are why the factors of the SITS are not presented in phases (or stages like the CASEL Implementation Rubric) on the SITS survey. In short, the cumulative effect of building and renewing the system of implementation over time is more critical than the specific implementation of one or two factors at a specific time.

In terms of the ecological levels, the SITS factors primarily map onto the school level. This was done for two reasons. First, most of the empirical literature specifying the network of factors supporting implementation exists at the school level. The district and community levels impact the school level, but the school level is where the decisions are made which support or hinder fidelity and sustainability. Second, few schools have the resources or the motivation to measure implementation at the classroom level when not forced to do so by outside forces (e.g., for a university research project), making it non-sustainable and non-scalable in practice. Thus, focusing on the school level maintains the most empirical support while having the strongest likelihood of leading to a science of implementation that is used under typical circumstances. All of the levels are important, but that does not necessitate that they all have the same practical utility in terms of designing measurements.

**Research questions and hypotheses.** Several research questions (hereafter labeled as “Qs”) and several hypotheses (hereafter labeled as “Hs”) will be examined to see how well the system of implementation, as measured by the SITS, is able to adhere to the conceptual underpinnings outlined above in this thesis. An analysis is labeled a research
question if the conceptualization suggests that there may be a relation between the system of implementation and another construct but the nature of the relation is not clear. An analysis is labeled a hypothesis if the conceptualization suggests a clear direction for the relation between the system of implementation and another construct.

**Q1.** Does the relationship between the system of implementation and other constructs differ depending upon who is evaluating the system of implementation? That is, do the DSACS Liaisons and the School Coordinator agree about the SITS scores for the schools? If there are differences in opinion, in what ways do the differing opinions affect the research questions and hypotheses to be elaborated below?

**Q2.** What aspects of the school structure affect the system of implementation? That is, (Q2a) does the grade configuration of the school (i.e., if it is an elementary school, middle school, or high school) affect the SITS scores? And, (Q2b) does the size of the school (i.e., school enrollment) affect the SITS scores? Grade configuration is quite uncertain. Presumably, smaller enrollment will predict higher SITS scores, but there are many other factors that may affect this relationship (e.g., student to faculty ratio, grade configuration, and financial resources).

**H1.** Schools with more structural risk factors (at the school, district or community level) will have weaker systems of implementation. That is, (H1a) schools with higher student to faculty ratios will have lower SITS scores, and (H1b) schools within districts and communities with lower socioeconomic status (SES) will have lower SITS scores.

**H2.** Schools with a better school environment will have higher implementation scores within a given year. That is, (H2a) higher school student climate will predict
higher SITS scores, (H2b) higher school staff climate will predict higher SITS scores, (H2c) higher student attendance will predict higher SITS scores, (H2d) higher staff mobility will predict lower SITS scores, (H2e) higher suspension rates will predict lower SITS scores, (H2f) higher bullying rates will predict lower SITS scores, and (H2g) higher rates of violence will predict lower SITS scores.

**H3.** The system of implementation will be more robust for schools in the continuing delivery phase than those in the preparation and initial delivery phases. That is, the longer a school has been implementing SECD, the higher its SITS score should be.

**H4.** Schools that receive more support from experts will establish their systems of implementation faster and change their school environment faster, especially during the preparation and initial delivery phases. That is, (H4a) schools that receive more assistance from experts will have higher SITS scores, especially if they are in the first year of implementation, (H4b) schools that receive more assistance from experts will achieve greater improvements in the school climate in the following year, and (H4c) the effect of expert support on climate change in the following year will be mediated by higher SITS scores. (Non-climate aspects of the school environment should follow the same pattern, but behavioral indicators represent more long term problems. This means that more than one year must pass between the building of the system of implementation and changes in these outcomes. The limited timeframe of the data for this thesis prohibits more extensive testing of this hypothesis.)

**Method**

**Setting**
The Schools Implementing Towards Sustainability (SITS) Scale was distributed to all active schools participating in the Developing Safe and Civil Schools (DSACS) Project. There were 150 active schools in 2008-2009 and 126 active schools in 2009-2010 (see Table 1). At the time that the SITS was completed, DSACS was an ongoing, longitudinal, action-research project aimed at improving school climate and academics and decreasing violence through implementing SECD in New Jersey schools (Elias, 2009). Action-research refers to a process in which the investigation is systematic, participatory, reflective, and cyclic. Action-research is “...not just research which we hope will be followed by action! It is action which is researched, changed and re-researched, within the research process by participants” (Wadsworth, 1998, “What Participatory Action Research is - and is not!” para. 4).

All participating DSACS schools agreed to certain responsibilities. They agreed to identify a SECD coordinator as the primary contact person for the DSACS Liaison staff member; to organize an SECD committee to plan, implement, and evaluate SECD in the school; to develop an annual and multiyear SECD implementation plan and to set 8-week cycle goals; and to attend DSACS professional development opportunities throughout the year. Thus, schools joining DSACS were expected to engage in many evidence-based implementation and sustainability practices.

The school SECD committee was presented with a menu of required DSACS services that they were expected to complete over a period of 3 years. Each school’s committee, however, was able to choose their own sequence to address the needs of that particular school. The menu included evaluating the school climate of the students and staff and tracking behavioral measures important to the school’s goals; articulating core themes
and values to connect school policy, curriculum and programs; reviewing the school’s
code of conduct, harassment and bullying policies; stating attendance policies as
statements of positive expectations; coordinating existing and future SECD and other
prevention and promotion programs; applying SECD to academic areas and classroom
management; and increasing student voice and engagement.

Participants

Participants include all of the DSACS Liaisons and School Coordinators who
completed a confidential implementation scale, the SITS (see Table 1 for survey counts).
Participants also include all students and teachers/staff that completed an anonymous
DSACS school climate survey (when that school has also completed the SITS). DSACS
obtained approval from the Institutional Review Board (IRB) of Rutgers, The State
University of New Jersey and the New Jersey Department of Education (NJ DOE) to
collect confidential data from DSACS Liaisons, School Coordinators, and schools and to
collect anonymous data from students, teachers, and staff as part of ongoing program
evaluation. All schools that have at least one SITS Scale (by a DSACS Liaison and/or
Coordinator) will be included in this paper. This results in 157 unique schools, when
looking at both respondent types (i.e., Liaison and Coordinators) for the SITS and both
time points (i.e., 2008-2009 and 2009-2010).

SITS demographics. All DSACS Liaisons and all School Coordinators were invited
to complete the SITS. The DSACS Liaison is the outside expert on SECD and the contact
person for the school. All DSACS Liaisons completed the SITS at Time 1 (n = 13) and at
Time 2 (n = 11). Two DSACS Liaisons were White males and one was an African
American female. The remaining DSACS Liaisons were White females. The median
The number of surveys completed by a DSASC Liaison was 4 at Time 1 (range = 1-36; only 5 people completed more than 4 surveys) and 3 at Time 2 (range = 1-42; only 4 people completed more than 5 surveys).

The School SECD Coordinator is the person in the school who leads the SECD effort in the school and acts as contact person within the school. There were no refusals at either time. Demographic information was not collected during the administration of the SITS, but most of the DSACS School Coordinators were White and the vast majority was female. Most School Coordinators were the contact person for only one school, but some were the coordinator for the entire district. The median number of surveys completed by a School Coordinator was 1 at both times (range = 1-6 at Time 1; range = 1-8 at Time 2; although this is based off of IP Addresses rather than names, so the numbers may be off if Coordinators switched or shared computers). All School Coordinators had other roles within the school; the most common roles were being a teacher or a school mental health professional (e.g., school psychologist).

**Climate demographics.** All students were invited to complete the student climate surveys; and all staff were invited to complete the staff climate surveys. There were no refusals by students or staff that were in attendance at the time of survey (see Table 1 for survey counts and Table 2 for climate demographics). Students usually completed the survey on a typical school day, and staff usually completed the survey during a faculty meeting. There were no provisions to gather data from any absentees.

**School demographics.** All schools are within the State of New Jersey (see Table 3 for school demographics). Almost all schools were public schools. A little under 1/5 were from school districts with low socio-economic status.
Over 1/3 were elementary schools, 1/4 were middle schools, and 1/16 were high schools – with the rest having another type of grade configuration. ES was defined as grade combinations of PK-4, PK-5, PK-6, K-4, K-5, K-6, 2-5, 3-5, 3-6, 4-5, 4-6 (with PK meaning pre-kindergarten and K meaning kindergarten). The most common grade combinations were K-5, PK-5, PK-6 and K-6 (in descending order). MS was defined as grade combinations of 5-8, 6-8, 7-8, with 6-8 being the most common combination. HS was defined as the combination of grades 9-12.

Measures

**SITS Time 1.** The SITS was created in June of 2009 by Dominic C. Moceri and Maurice J. Elias of the Social-Emotional Learning (SEL) Lab at Rutgers, The State University of New Jersey after a comprehensive review of the research literature on evidence-based SECD implementation and sustainability strategies. Duplicate strategies and those that the DSACS Liaisons or Coordinators were not likely to know were dropped. A rough draft of the SITS was given to the two primary DSACS Liaisons (i.e., the Director of Field Services and the Project Director), who provided feedback about the format and which questions contained information that DSACS Liaisons and School Coordinators would likely be able to rate reliably.

The final survey was approved by the DSACS Director of Field Services before being distributed to all of the DSACS Liaisons. In the end only 15 true/false implementation items remained, which was determined to be a reasonable balance of thoroughness and sustainability (i.e., not creating a survey so lengthy or demanding that few, if any, DSACS Liaisons or School Coordinators would complete it) (see Appendix A for the items in the SITS survey). The DSACS Liaisons completed the survey between June and
August of 2009 for the 2008-2009 academic year (i.e., Year 4 of DSACS). All but 5 School Coordinators completed the survey between September and October of 2009, with the last five completing the survey by November of 2009, for the 2008-2009 academic year.

**SITS Time 2.** The wording of the 15 questions from the SITS for Time 1 was kept the same, but the response options to the questions were changed. Based upon feedback from the DSACS Liaisons as well as a review of psychometrics, the true/false options were changed into a 4-point scale, which ranged from “False, not at all true” to “Very true.” While this change means that the SITS at Time 1 and the SITS at Time 2 are not identical and do not fit into a traditional psychometrics framework, the premise of modifying measurement scales fits with the action-research framework on which DSACS is premised. The DSACS Liaisons completed the SITS for Time 2 between June and July of 2010 for the 2009-2010 academic year (i.e., Year 5 of DSACS). All but one School Coordinator completed the SITS for Time 2 between April and July of 2010, with the last one completing the survey in August.

**Climate survey.** Every school participating in DSACS is strongly encouraged to participate in a school climate survey, which is completed by the students and staff on a yearly basis. The student climate survey, known as the DSACS Climate Survey-Student version (DSACS-CS-Stu) (Elias, 2009), consists of 20 questions (not counting demographics questions) (see Appendix E). The questions consist of an adapted version of the School as A Caring Community Profile-II (SCCP-II; Lickona & Davidson, 2004) as well as additional questions to measure the level of student and staff approval, the utility of learning, and student pride. The staff climate survey, known as the DSACS
Climate Survey-Staff version (DSACS-CS-Staff) (Elias, 2009) consists of the same 20 questions from the student climate survey plus 20 additional climate questions (see Appendix F). The questions were derived from the same sources as the student version plus the addition of the School Organisational Health questionnaire (Hart, et al., 2000). Both climate surveys are the same for Time 1 (2008-2009) and Time 2 (2009-2010).

**Bullying survey.** The end of the climate survey for the students contains some general bullying questions. These are considered a separate measure from the school student climate measure. The only bullying item being considered for this thesis is “Students are often bullied or teased in my school.”

**DSACS variables.** Two variables (cohort & tier) from the DSACS Initiative will be utilized in analyses.


**Tier.** The tier refers to the format of DSACS services that a school was assigned. All tiers had an initial DSACS Liaison meeting for climate data review and initial planning, but the format of DSACS services after that varied based upon the assigned tier level. Cohorts 1-3 were assigned to Tier based upon their perceived need by the DSACS staff. Cohort 4 was randomly assigned to Tiers in a stratified manner (based upon school type and size). For all cohorts, all schools within the same district were assigned the same Tier.
Tier 1 schools received a DSACS e-mail and telephone buddy. This level of support included 1 in-person meeting. Continued support was delivered via computer or phone only.

Tier 2 schools received process focused, on-site support. They received everything Tier 1 did plus limited on-site visits to help set up the infrastructure for DSACS implementation (i.e., creating a committee structure, prioritizing goals). However, they did not receive any school-based training from DSACS. The DSACS Liaison had a consultation and capacity-building role.

Tier 3 schools received process and content on-site support. They received everything Tiers 1 and 2 did plus they received content-level expertise (i.e., direct help integrating SECD with academics, assistance with creating a core school theme, assistance with professional development training to school staff, and a meeting with the DSACS Liaisons to help them interpret and plan for climate survey feedback).

**NJDOE variables.** Several variables from the New Jersey Department of Education (NJDOE) will be utilized in analyses. All of variables were publically available through the NJDOE School Report Cards online database (http://education.state.nj.us/rc/) or the NJDOE Electronic Violence and Vandalism Reporting System (EVVRS) (https://homeroom2.state.nj.us/DOE_EVVRS/jsp/login.jsp). NJDOE variables include grade configuration (i.e., what grades are in the school), school size (i.e., enrollment), student to faculty ratio, staff mobility, student attendance rate, number of suspensions, school violence rates, and the District Factor Grouping (DFG). DFG is an indicator of socio-economic status (SES); and is determined with each Decennial Census based upon the following 6 variables: 1) Percent of adults with no high school diploma, 2) Percent of
Reliability of the SITS

**Internal reliability.** The internal consistency reliability (Cronbach, 1951) of the SITS was good at Time 1 for DSACS Liaisons ($\alpha = .87$) and for School Coordinators ($\alpha = .85$) and very good at Time 2 for DSACS Liaisons ($\alpha = .96$ for primary coding, $\alpha = .92$ for alternative coding) and School Coordinators ($\alpha = .93$ for primary coding, $\alpha = .89$ for alternative coding), using the conventions from Nunnally & Bernstein (1994).

**Scale creation.** The SITS scale (SITS-Tot) was computed by summing the scores of the true statements (see Appendix B for details on how SITS-Tot is calculated for each time period). Specific subscales were neither hypothesized nor tested for at this stage of measure development as the order of implementation was highly flexible within the DSACS framework and any theoretical subscales might change over time.

**Parallel analyses (Q1).** The correlation between the primary and alternative coding of the SITS at Time 2 was very high for DSACS Liaisons, $r(118) = .91, p < .000$, and high for School Coordinators, $r(61) = .77, p < .000$. Therefore, only the results for the primary coding of Time 2 will be provided in subsequent analyses unless the results for the alternative coding were substantially different.

A paired-samples t-test revealed a significant difference between the SITS scores by the DSACS Liaisons and the School Coordinators at Time 1 ($M_L = 8.99, SD_L = 3.81; M_C = 10.14, SD_C = 3.91$, where $L =$ Liaison and $C =$ Coordinator), $t(68) = -2.17, p < .05$, and...
at Time 2 (\(M_L = 23.28, SD_L = 14.64; M_C = 27.72, SD_C = 10.81\)), \(t(60) = -2.99, p < .01\). At each time point the School Coordinators provided higher SITS scores.

The differences in scores at both time points indicate the need for parallel analyses of the research questions and hypotheses for all preceding and proceeding analyses even though there was a medium correlation between the two SITS scores at Time 1, \(r(67) = .34, p < .01\), and a large correlation in Time 2, \(r(59) = .62, p < .000\), at Time 2, using the conventions from Cohen (1992), where an \(r\) of .1 is small, .3 is medium and .5 is large. When the alternative coding for Time 2 was used, the correlation was medium, \(r(59) = .31, p < .05\). As each respondent has a unique viewpoint (i.e., the Liaison as the expert on SECD and the Coordinator as the expert on the school), neither group’s responses will be posited to be more true than another.

**Stability over time.** There was a strong relationship between the SITS scores at Time 1 and 2 for the DSACS Liaisons, \(r(99) = .57, p < .000\), and for the School Coordinators, \(r(32) = .60, p < .000\). The high stability over time approaches the traditional test-retest lower bound value of .667 (Nunnally & Bernstein, 1994), which is remarkable as a) sustainability is an ongoing process that must be actively maintained rather than a permanent characteristic, b) 10-13 months passed between Time 1 and Time 2 for the DSACS Liaisons and 5-10 months passed between Time 1 and Time 2 for the School Coordinators, and c) the response choices changed.

When using the alternative coding system for Time 2, paired-samples t-test revealed a significant increase between Time 1 and Time 2 for SITS scores by DSACS Liaisons (\(M_1 = 8.77, SD_1 = 4.14; M_2 = 10.15, SD_2 = 4.43\), where 1 = Time 1 and 2 = Time 2), \(t(100) = \)
3.37, \( p < .001 \), and SITS scores by School Coordinators (\( M_1 = 10.15, SD_1 = 3.93; M_2 = 12.67, SD_2 = 3.51 \)), \( t(33) = 4.39, p < .000 \).

Two different methods were utilized to determine how many schools showed a meaningful change in their SITS scores across time. The first used the reliable change index (RCI; Jacobson & Truax, 1991), which determines whether or not change for a given case is statistically reliable. Part of the RCI relies on the short-term test-retest reliability of the measure. As this is unknown for the SITS, the RCI was calculated twice using two different assumptions. The first time it was assumed that the short-term test-reliability was equal to the above 6-12-month test-retest reliability and the second time it was assumed that it was equal to .80, a good short-term test-retest reliability. The second method of calculating how many schools showed a change between Time 1 and Time 2 looked at how many schools’ SITS scores changed by half a standard deviation or more. When using the alternative coding system, 6.9-38.6% of the schools increased in their DSACS Liaison SITS scores and 8.8-50.0% of the school increased in their School Coordinator SITS scores (see Table 4 for details).

**Summary of reliability.** In summary, the SITS has good reliability. Internal reliability is high. There was a significant association between the implementation and sustainability ratings (i.e., the SITS) by outside experts (i.e., DSACS Liaisons) and by inside experts (i.e., School Coordinators); experts within the school provided higher ratings. There was a significant increase across time in the implementation and sustainability ratings, but the ratings were relatively stable across time.

**Concurrent Validity**
Preliminary frequency analyses revealed that the distribution of tier, cohort and grade configuration for the SITS surveys was heavily unbalanced. Therefore, the present data prevent the examination of any interaction effects: tier, cohort and grade configuration will be looked at individually.

**Implementation assistance by experts (H4a).** All cohorts and grade configurations were analyzed together for the first set of tier analyses, which indicate the level of expert assistance for implementation. One-way analysis of variances (ANOVAs) of Time 1 revealed a medium effect for Tier for DSACS Liaisons, $F(2, 134) = 14.72, p < .000$, partial eta squared ($\eta^2$) = .18, and a medium effect for Tier for School Liaisons, $F(2, 68) = 5.69, p < .01$, partial $\eta^2 = .14$, using the conservative guidelines of Cohen, Cohen, West, and Aiken (2002), where .02 is small, .13 is medium and .26 is large. Post-hoc analyses of Time 1 using Tukey's HSD (Honestly Significant Difference) revealed that all Tiers were different from each other in the hypothesize manner. For DSACS Liaisons Tier 1 ($M = 5.13, SD = 3.86, n = 30$) was less than Tier 2 ($M = 8.41, SD = 3.70, n = 66$) and both were less than Tier 3 ($M = 10.00, SD = 3.81, n = 41$) (i.e., Tier 1 < Tier 2 < Tier 3) (see Figure 1). For School Coordinators Tier 1 ($M = 7.40, SD = 4.12, n = 15$) was less than Tier 2 ($M = 11.24, SD = 3.63, n = 33$) and Tier 3 ($M = 10.35, SD = 3.43, n = 23$); but Tiers 2 and 3 did not differ from each other (i.e., Tier 1 < Tier 2 = Tier 3).

One-way ANOVAs of Time 2 revealed a small effect for Tier for DSACS Liaisons, $F(2, 115) = 5.01, p < .01$, partial $\eta^2 = .08$, but no effect for School Coordinators, $F(2, 62) = 1.82, n.s.$ Post-hoc analyses of Time 2 using Tukey's HSD revealed that Tier 1 ($M = 14.32, SD = 13.90, n = 19$) was less than Tier 3 ($M = 26.05, SD = 13.17, n = 41$) for DSACS Liaisons; Tier 2 ($M = 21.28, SD = 13.55, n = 58$) was not different from the
other tiers (i.e., Tier 1 < Tier 3). The only difference using the alternative coding for Time 2 was that Tier 1 was also less than Tier 2 for DSACS Liaisons.

The above analyses were rerun with only Cohort 4, as tier was randomly assigned for Cohort 4. The ANOVA of Time 1 revealed a large effect for Tier for DSACS Liaisons, $F(2, 41) = 7.11, p < .01$, partial $\eta^2 = .26$. Post-hoc analyses of Time 1 using Tukey’s HSD revealed that Tier 1 (M = 4.58, SD = 3.73, n = 19) was lower than Tier 2 (M = 8.90, SD = 3.60, n = 10) and Tier 3 (M = 8.67, SD = 3.56, n = 15) for DSACS Liaisons (i.e., Tier 1 < Tier 2 = Tier 3). At Time 2 tier was not significant for DSACS Liaisons using the primary coding; but tier had a medium effect using the alternative coding, $F(2, 39) = 4.12, p < .05$, partial $\eta^2 = .17$, with Tier 1 (M = 6.33, SD = 4.01, n = 12) scoring less than Tier 3 (M = 10.76, SD = 3.85, n = 17); Tier 2 (M = 9.08, SD = 4.42, n = 13) was not different from the other tiers. Tier was not significant for School Coordinators at either time point; however, it should be noted that the sample was small (n_1 = 22, n_2 = 21) for Cohort 4. In short, Tier was persistently related to SITS scores, especially for DSACS Liaisons, and Tier 3 schools persistently scored higher on the SITS than Tier 1 schools.

**Implementation duration (H3).** All tiers and grade configurations were analyzed together for the first set of cohort analyses, which indicate how long a school has been implementing SECD through DSACS. One-way ANOVAs of Time 1 revealed a small effect for Cohort for DSACS Liaisons, $F(3, 133) = 4.88, p < .00$, partial $\eta^2 = .10$. Post-hoc analyses of Time 1 using Tukey’s HSD revealed that Cohort 3 (M = 9.66, SD = 3.84, n = 58) was greater than Cohort 2 (M = 7.00, SD = 4.10, n = 27) and Cohort 4 (6.95, SD = 4.13, n = 44) (i.e., Cohort 3 > Cohort 2 = Cohort 4); Cohort 1 (M = 8.00, SD = 3.34, n = 8) was not different from the other cohorts. No other cohort ANOVAS were significant.
However, when the above analyses were rerun with only Tier 2, no cohort ANOVAS were significant. The above analyses were rerun with only Tier 2 because chi-square analyses revealed that the Tiers were not evenly distributed among the Cohorts for Time 1, \( \chi^2 (6, N = 139) = 32.33, p < .000 \), or for Time 2, \( \chi^2 (6, N = 123) = 25.35, p < .000 \); and very few schools had Tiers 1 or 3 in Cohorts 1 and 2. This suggests that Tier was a confounding variable in the preceding paragraph and only significant results due to Cohort are really due to the Tier distribution in this dataset.

**Summary of concurrent validity.** In summary, schools that received more support from outside experts (i.e., having a higher Tier) utilized more evidence-based implementation and sustainability strategies (i.e., scored higher on the SITS) than schools that received less support. The implementation duration (i.e., Cohort) was not related a higher use of these strategies (i.e., higher SITS scores) when the level of support from outside experts (i.e., Tier) was kept equal.

**School Structure Analyses (Q2)**

All tiers and cohorts were analyzed together for the first set of the school structure analyses, which includes grade configuration and school size.

**Grade configuration (Q2a).** Grade configuration refers to a school being an elementary school (ES), middle school (MS), or high school (HS). One-way ANOVAs of Time 1 revealed no significant effect for grade configuration for DSACS Liaisons, \( F(2, 105) = 1.56, n.s. \), or for School Coordinators, \( F(2, 54) = 0.27, n.s. \). At Time 2 violations of Levene’s test of equality of variance required independent t-tests for each comparison, using .0167 (.05/3) for significance and .0333 (.1/3) for trend. For DSACS Liaisons at Time 2 there was a trend for ES (M = 24.61, SD = 11.47, n = 44) scoring higher than HS (M = 15.47, SD = 14.77, n = 17), \( t(59) = 2.57, p = .013 \); MS (M = 23.61, SD = 15.51, n = 33) was not different from the other grade configurations (i.e., ES > HS). Grade
configuration was not significant for School Coordinators at Time 2. Chi-square results for grade configuration and Tier were not significant for Time 1 or 2.

The above analyses were then rerun with only Cohort 4 because a) grade configuration was stratified in Cohort 4 and b) chi-square results for grade configuration and Cohort were significant at Time 1, \(\chi^2 (6, N = 106) = 15.51, p < .05\), and at Time 2, \(\chi^2 (6, N = 97) = 18.32, p < .01\). There was a medium effect for grade configuration for DSACS Liaisons at Time 1, \(F(2, 25) = 3.50, p < .05\), partial \(\eta^2 = .24\), with ES (M = 11.14, SD = 2.27, n = 7) showing a trend of scoring higher than MS (M = 6.70, SD = 4.42, n = 10) and HS (M = 6.38, SD = 4.27, n = 8) (i.e., ES > MS = HS). No other analyses were significant; although most had low power due to small samples.

**School size (Q2b).** The school’s total enrollment did not significantly correlate with DSACS Liaisons or School Coordinator SITS scores at Time 1. Enrollment data is unavailable for Time 2; but since the total enrollment between Time 1 (i.e., 2008-2009) and Time 0 (i.e., 2007-2008) is nearly identical, \(r(111) = .998, p < .000\), and total enrollment between Time 1 and Time 3 (i.e., 2004-2005) is also nearly identical, \(r(110) = .97, p < .000\), it is safe to use the Time 1 enrollment for the Time 2 analyses. There was a significant correlation for DSACS Liaisons at Time 2, \(r(107) = -.25, p < .01\), but not for School Coordinators at Time 2. When two outliers from the DSACS Liaisons at Time 2 correlation were removed, the relationship was stronger, \(r(105) = -.34, p < .000\).

**Summary of structural analyses.** In summary, elementary schools usually scored higher than high schools on implementation and sustainability (i.e., the SITS); but this finding was limited to ratings done by expert outsiders (i.e., DSACS Liaisons). Smaller schools scored higher than larger schools on implementation and sustainability (i.e., the
Construct Validity (H1, H2)

All tiers, cohorts, and school structure analyses were analyzed together while examining the school structural risk factors and the school environment.

Structural risk factors (H1). Low DFG schools (M = 5.93, SD = 3.34, n = 27) scored lower than non-low DFG schools (M = 8.95, SD = 4.15, n = 102) on the DSACS Liaison SITS scores at Time 1, \( t(127) = -3.35, p < .001 \). Low DFG schools (M = 14.41, SD = 9.70, n = 17) scored lower than non-low DFG schools (M = 23.87, SD = 14.04, n = 96) on the DSACS Liaison SITS scores at Time 2, \( t(29.40) = -3.43, p < .005 \). The alternative coding for Time 2 for DSACS Liaisons was not significantly related to DFG status. School Coordinator SITS scores were not significantly related to DFG status. Chi-Square tests for DFG and Tier and chi-square results for DFG and Grade Configuration were not significant at Time 1 or Time 2, which suggests that the DFG results are independent of those two factors. The student to faculty ratio was not significantly correlated with the DSACS Liaison or School Coordinator SITS at Time 1; it is unavailable for Time 2.

School environment (H2). Student Climate showed a trend of an association with the alternative coding for DSACS Liaisons SITS scores at Time 2, \( r(42) = .27, p < .10 \). No other Student Climate correlations were significant.

The full 40-item Staff Climate survey had a medium association with the School Coordinator SITS score at Time 1, \( r(36) = .48, p < .01 \). The 20-item Staff Climate survey that duplicates the Student Climate survey had a small trend of an association with the
DSACS Liaison SITS scores at Time 1, $r(77) = .22, p < .10$, and a medium association with the School Coordinator SITS scores at Time 1, $r(36) = .43, p < .01$. No other Staff Climate correlations were significant.

Bullying rates were associated with DSACS Liaison SITS scores at Time 1, $r(75) = -.30, p < .01$; but were not associated with School Coordinator SITS scores at Time 1. Bullying was not associated with either SITS score at Time 2.

Student attendance rates showed a trend of an association with the DSACS Liaison SITS score at Time 1, $r(124) = .16, p < .10$; after removing 6 outliers the relationship was stronger, $r(118) = .24, p < .01$. Student attendance rates were not associated with the School Coordinator SITS at Time 1.

Staff mobility rates were not associated with the DSACS Liaisons or School Coordinator SITS scores at Time 1; but, after removing 5 outliers, staff mobility was positively associated with DSACS Liaison SITS scores, $r(119) = .18, p < .05$, which is contrary to the hypothesis. That is, for Time 1 greater staff mobility was associated with higher DSACS Liaison SITS scores rather than lower scores.

Higher student suspensions were associated with lower DSACS Liaisons SITS scores at Time 1, $r(124) = -.19, p < .05$, but not lower School Coordinator SITS scores. After removing 12 outliers higher suspensions showed a trend of association with lower School Coordinator SITS scores at Time 1, $r(59) = -.23, p < .10$, but not DSACS Liaison SITS scores. However, after removing 13 outliers higher suspensions was not associated with lower DSACS Liaisons or School Coordinator SITS scores.
Violence rates were not associated with DSACS Liaison or School Coordinator SITS scores at Time 1. It was not possible to look at student attendance, staff mobility, student suspensions, or violence rates at Time 2 because the data is unavailable.

**Summary of construct validity.** In summary, both the school structure and the school environment were associated with implementation and sustainability (i.e., the SITS); however, these findings were primarily related to implementation and sustainability as rated by outside experts (i.e., DSACS Liaisons). Consistent with the hypotheses, better interpersonal relationships within the school (i.e., school climate) and higher student attendance were associated with higher implementation and sustainability rates; whereas lower socio-economic status (i.e., DFG) and lower bullying were associated with lower implementation and sustainability rates. Lower suspension rates were sometimes associated with lower implementation and sustainability. Higher staff mobility was associated with higher implementation and sustainability, which was contrary to the hypothesis. The only aspects of school structure and school environment not associated with implementation and sustainability were the school’s student to faculty ratio and the school’s violence rates by students.

**Predictive Validity (H4b, H4c)**

Unfortunately, the sample size was small for all predictive validity analyses. There are several reasons for this. One, about half of the schools at Time 1 had a student climate survey; and about half of the schools at Time 1 had a staff climate survey. Two, less than half of the schools at Time 2 had a student climate survey; and only about one-quarter of the schools at Time 2 had a staff climate survey. Three, a little over one-quarter of the schools that were active at Time 1 were not active at Time 2. Four, while almost all
active schools have a DSACS Liaison SITS survey, only about half of those schools have a Coordinator SITS survey. Therefore, predictive validity analyses would be expected to be only one-quarter of the total number of student climate and bullying surveys for DSACS Liaisons, one-eighth of the total number of student climate and bullying surveys for School Coordinators, one-twentieth of the total number of staff climate surveys for DSACS Liaisons, and one-tenth of the total number of staff climate surveys for School Coordinators. Those percentages correspond to the actual numbers in the below analyses.

Changes in Climate scores were calculated by subtracting the Time 1 scores from the alternative coding of Time 2; and changes in Bullying rates were calculated by subtracting Time 1 scores from Time 2 scores. DSACS Liaison SITS scores at Time 1 strongly predicted changes in Staff Climate scores when using the 20-item version of the survey that duplicates the Student Climate survey, $r(8) = .64, p < .05$; but not when using the 40-item version of the Staff Climate survey. DSACS Liaison SITS scores at Time 1 did not predict changes in Student Climate scores or Bullying rates. School Coordinator rates did not predict changes in Student Climate, either version of the Staff Climate survey, or Bullying rates. One reason for the null results for changes in Student Climate is that there was an extremely high correlation between Time 1 and Time 2, $r(17) = .94, p < .000$; t-tests revealed that Student Climate did not change over time. Similarly, for Bullying there was a high correlation between Time 1 and Time 2, $r(16) = .81, p < .000$; t-tests revealed a trend for Bullying to increase between Time 1 ($M_1 = 3.06, SD_1 = .44$) and Time 2 ($M_2 = 3.18, SD_2 = .37$), $t(17) = 1.80, p < .10$.

More complicated (i.e., mediator) analyses of the SITS at Time 1 predicting changes in outcomes were not attempted for three reasons. One, tier was unable to predict changes
in Student or Staff Climate (40-item or 20-item version) scores. Two, the sample sizes for having two climate surveys of the same type and two SITS scores of the same type were very small (i.e., the largest sample was composed of 19 schools). No other outcome variables are available for Time 2.

**Summary of predictive validity.** In summary, higher implementation and sustainability (i.e., the SITS) at Time 1 were associated with more improvement between Time 1 and Time 2 for the staff school climate. Implementation and sustainability at Time 1 was not associated with changes in either the student school climate or the overall bullying rates between Time 1 and Time 2.

**Discussion**

Overall, the “Schools Implementing Towards Sustainability” (SITS) scale has good reliability as well as good concurrent and construct validity for measuring evidence-based strategies of implementation and sustainability. Higher SITS scores were associated with greater support from outside aspects and these types of school structures: being an elementary school, being a smaller school, and not belonging to a low socio-economic status district. Higher SITS scores were also associated with better staff school climate, less bullying, and higher student attendance. While predictive validity of the SITS is promising, with higher SITS scores predicting greater positive changes in staff climate, a small sample size limits the generalizability of these findings. Although the reliability findings apply to both the ratings done by outside experts and by inside experts, the validity analyses usually apply only to the ratings done by outside experts.

**Comparison with existing literature.** The SITS is highly consistent with the literature on implementation and sustainability of school-based prevention and promotion
interventions (SBPPI). First, ecological levels matter, as argued by Greenberg et al. (2005) and others, which is demonstrated in this study by schools in lower socio-economic status districts scoring lower on sustainability (i.e., the SITS). Second, elementary schools are associated with higher implementation rates, which is consistent with CASEL (2005) identifying more “Select SEL” programs for elementary schools. Third, implementation and sustainability strategies described by many researchers (e.g., Berkowitz, & Bier, 2006; CASEL Safe & Sound, 2005; Commins & Elias, 1991; Elias, 2007; Greenberg et al., 2005; Ji et al., 2008) were consistently associated with better school environments. For example, positive staff school climate was associated with higher implementation, as stated by King and Newmann (2000).

Finally, the SITS is consistent with the literature on measuring implementation and sustainability of SBPPI. Like the SITS, the CASEL Rubric (Ji et al., 2008) found improvement between its first administration to the participating schools and the second year’s administration (Peter Ji, personal communication, August 21, 2010). However, given the differences in how the two measures are intended to be used, organized and assessed (see “SITS versus CASEL Rubric” under “The Present Study” above and Appendix C), other empirical comparisons between the two measures are not currently possible.

Nevertheless, there were three potential discrepancies between the research literature and the above results for the SITS. One, schools in cohorts with more years in the DSACS project, which presumably were in higher phases of implementation, did not have higher implementation scores than the schools in more recent cohorts. Higher phases of implementation usually correspond to greater implementation and sustainability
Two, higher student to faculty ratio was not associated with lower implementation scores. Previous work identifies student ratios as risk factors for problematic behavior (SAMSHA, 2002). Three, higher staff mobility (i.e., turnover) was associated with higher implementation scores rather than lower implementation. Turnover is usually considered a threat to implementation and sustainability (Elias, 2007).

**Reasons and implications.** Most findings were consistent with the conceptualization of implementation by both this study and Greenberg and colleagues (2005). The ecological levels of a school impact its implementation because implementation always occurs in a specific temporal, situational, and developmental context. A better school environment means that the school has the capacity to roll out the implementation faster. For instance, when staff school climate is positive, they can trust that the latest intervention by the administration is serious rather than a fad that will soon disappear. Whereas, when staff school climate is negative, it is probably due to a lack of trust with the administration and/or being overburdened, both of which make deploying a new intervention less appealing. Meanwhile, certain school structures seem to enhance implementation. For instance, smaller schools may allow information within the school to disseminate more rapidly as training and monitoring encounter fewer hurdles. And, the ecological levels surrounding the school, such as the district’s socio-economic status, can restrict (or enhance) how financial resources can be used to create the training and structure for the intervention, especially in the first few years.

Nevertheless, the potential discrepancies between the research literature and the findings of this study must be addressed. Cohorts involved with SBPPI for a longer
duration may not have been associated with higher implementation scores for multiple reasons. The SITS was tested through a highly heterogeneous, process-oriented intervention that attempted to meet schools where they were and to address a given school’s specific needs, rather than have schools implement a rigid, pre-packed intervention where everyone received the same treatment. That is, DSACS emphasizes adaptability rather than manualization. The trajectory of intervention components for process-oriented interventions is more variable than manualized interventions, especially over shorter time periods, which makes cohort analyses less predictable. Therefore, the presumption that more experienced cohorts were more likely to be in the continuing delivery phase (or even substantially further in the initial delivery phase) was not necessarily accurate.

In addition, the SITS was not created and administered until the end of year that Cohort 4 joined DSACS. That means that all 4 Cohorts were exposed to the SITS at the same time. If the SITS had been a part of DSACS since its beginning, the Cohorts 1, 2 and 3 may have used it as an evaluation and planning tool before Cohort 4 rather than at the same time as Cohort 4. Similarly, DSACS and school personnel may have become more aware of sustainability as the project entered its final two years and emphasized it more than in the first few years. Dalton and colleagues (2006) report that it takes 3-5 years for positive changes to be noticeable and stable after implementing a program, so it is possible that all four cohorts are in the same phase (e.g., initial delivery phase), which would mean that they should have the same SITS scores.

Other discrepancies may have reasons for their existence, too. For instance, the student to faculty ratio may not have been associated with the SITS because this study
used the student to faculty ratio rather than the student to teacher ratio (which was
reported by SAMSHA, 2002); and some schools may have received a boost to their ratio
by non-teacher staff. Meanwhile, there are two possible reasons for why higher staff
mobility was associated with higher SITS scores. First, even though higher staff mobility
requires additional training in the short-term, this could still be a good thing for the long-
term sustainability of the intervention if the person entering the school is supportive of
the intervention or the person leaving the school is unsupportive. In fact, decisions that
“favor hiring and retaining staff who are compatible with goals of program” is a key
sustainability strategy (Elias, 2007, p. 91).

Alternatively, higher staff mobility may have been a problem that spurred the school
to make more dramatic changes. While the conceptualization of this thesis argues that
problems in the school usually hinder implementation and sustainability efforts,
sometimes highly visible problems (such as staff mobility) demonstrate the need for
change with great clarity. This can motivate schools to not only try an intervention (such
as SECD) but also to implement the intervention with greater urgency and consistency.

While not a discrepancy, the SITS only demonstrated some predictive validity in this
study: higher SITS scores at Time 1 predicted greater improvements in the staff school
climate but not improvements in student school climate or bullying. Given the very small
sample and the short time frame though, this finding is encouraging and unsurprising.
The DSACS theoretical model states that changes in staff school climate leads to changes
in student school climate, which leads to changes in student behavior, which leads to
changes in student academic achievement (M. J. Elias, personal communication,
February 5, 2009). Therefore, one would not expect student climate to change so soon in the process.

Finally, it is useful to consider why there were fewer significant findings for school-based Coordinators than for the external Liaisons. The simplest explanation is that implementation and sustainability must be evaluated (at least in part) by outside experts rather than inside experts. Outside experts have more training and more vantage points whereas inside experts may be unable to see the water in which they are swimming. Inside experts may lack the knowledge to rate their own school objectively. An alternative explanation is that outside experts are biased by irrelevant information. An outside expert, for instance, may confuse positive staff school climate with solid sustainability strategies.

This thesis, however, rejects both of those explanations in favor of a third explanation: outside experts possess expertise which makes them more qualified to compare different schools whereas inside experts possess expertise which makes them more qualified to know the details about what is happening in a school. Both were able to achieve high levels of internal reliability and rated their schools consistently across time. And, both were able to see changes over time. Thus, to compare implementation and sustainability across schools, one needs to rely on outside experts. But, if one simply wants to track changes within a school both outside and inside experts are reliable and valuable sources of information.

**Future research.** A great deal of future research is needed, especially with regard to the predictive validity of the SITS (and other related measures). The true test of a sustainability measure is not to be able to distinguish between schools but to be able to
predict who will still be implementing the intervention several years later (i.e., predictive validity). There are two important ways that this can be done with this sample. The first way is to revisit the schools three to five years after the first administration of the SITS (i.e., in 2012 to 2014) to see who is still implementing SECD and with what degree of quality and documented impact. The second way is to follow-up on student and staff behavior three to seven years after the administration of the SITS (i.e., in 2013 to 2016) to evaluate the improvements in outcomes at the schools.

As stated previously, the positive effects of SECD interventions are not visible and stable until 3-5 years have passed (Dalton et al., 2006). Therefore, testing predictive validity too early could be highly misleading. Moreover, the theory of sequential change for DSACS is that changes in staff perceptions of school climate lead to changes in student school climate, which lead to changes in student behaviors, which lead to changes in academics and standardized test scores. Clear theories of change are important so that benchmarks can be established. Benchmarks allow one that one is on or off track. If a school is not showing changes in its staff or student school climate by the second year of implementation, it is probably not going to see changes in its student behaviors or test scores in year three. Without theoretically-derived, observable benchmarks, useful interventions (e.g., DSACS) and measures (e.g., SITS) might be abandoned prematurely, either through leaders/stakeholders losing interest or governments/agencies removing funding too early.

The second most important task is to see what can be done to increase the validity of the SITS when completed by inside experts. One possibility is to expand the number of items, reword them, and/or provide better guidance about what constitutes a specific
value rating. Another possibility is to change how the SITS is intended to be used. The CASEL Rubric suggests that multiple staff within the school complete the measure and compare the results. In fact, the simple process of having multiple people complete it may even strengthen the sustainability of the intervention by bringing the key stakeholders within the school together. Of course, if future studies demonstrate that inside experts possess predictive validity by themselves, then this line of inquiry may not be needed. Future studies should seek to have the SITS completed by the entire SECD-related committee or team, in addition to the school-based Coordinator.

Another important task is to test the SITS in other settings and against other measures. Can the SITS demonstrate reliability and validity with school-based interventions that are not SECD? The conceptualization and the name of the SITS both suggest that it can be used for a wide variety of prevention and promotion programs that are based in schools, but it has only been tested with one (heterogeneous) SECD sample. How does the SITS compare to the CASEL Rubric when both are administered at the same setting? How does the SITS compare to fidelity measures that have been developed for specific SECD interventions? Can the administration of a generic implementation and sustainability measure (i.e., the SITS) provide additional useful information to a fidelity measure for a specific program?

Finally, the individual items and response choices on the SITS still need to be optimized. The potential items on the SITS were created based upon an extensive literature review, but the final items were chosen based upon the opinions and insight of key DSACS personnel. Different items might increase the validity of the SITS, especially if the items were selected through statistical analyses; but one must be careful about
adding more items because at some point greater validity comes at the cost of reduced scalability. Likewise, one must be careful in removing items. Some items may be endorsed rarely and, therefore, have low correlations with other items and decrease internal reliability; yet these same items may have significant predictive value when they occur. Likewise, some SITS items may be important for benchmarking in a developmental way, even if they are not strongest psychometrically. Therefore, decisions about keeping items should be made with reference to implementation and sustainability theory as well as psychometrics.

In terms of the two scale choices, this study suggest that the 4-point scale (i.e., false to very true) is more useful than the 2-point scale (i.e., true/false) for one primary reason: the correlation between the DSACS Liaisons (i.e., the outside experts) and the School Coordinators (i.e., the inside experts) was large when using the 4-point coding of Time 2; whereas it was medium when using the 2-point coding (for both Time 1 and Time 2). Greater similarity in ratings is more likely to encourage collaboration and foster dialogue the focuses on how to improve implementation and sustainability (rather than dialogue that disputes the rating of an individual item). Of course, only random assignment of the different scales by a future study can conclusively determine if this is true.

**Implications for practice.** The most important implication of this study is that implementation and sustainability can be measured and quantified in an affordable and scalable way (i.e., by using the SITS). The second most important implication is that this measurement can be used to track changes within a school, even within a one-year period. Finally, outside experts must be a part of the implementation and sustainability evaluation process to order to ensure validity, and it is suggested that a wider range of
internal experts complete the SITS (something that can also be the focus of future research).

The SITS is short, requires few resources, and requires no formal training. But most importantly, the SITS provides valuable information about the implementation and sustainability strategies being used by schools for prevention and promotion interventions. Researchers, consultants, and school personnel are highly encouraged to use the SITS to provide an objective assessment of the potential for sustainability of their interventions.
References


List of Appendixes

Appendix A. Schools Implementing Towards Sustainability (SITS) Scale .................. 60

Appendix B. Coding System for the SITS ............................................................... 63

Appendix C. Comparison of the CASEL Implementation Rubric and the SITS Scale ... 65

Appendix D. Mapping the Phases and Levels onto the SITS Scale ......................... 68

Appendix E. Student Climate Survey ................................................................. 70

Appendix F. Staff Climate Survey ..................................................................... 72
Appendix A
Schools Implementing Towards Sustainability (SITS) Scale

Answer Choices:
At Time 1 participants chose “True” [1], “False” [0], or “Don’t Know” [0] for each item.
At Time 2 participants chose “False, not at all true” [0], “Somewhat true” [1], “True” [2],
“Very true” [3], or “Don’t Know” [0] for each item.

Instructions:
This section will ask about how well the school has implemented social-emotional
character development (SECD). If the school is addressing SECD in more than one way,
please evaluate the school's SECD implementation as a whole.

Please, select whether the following statements are true or false. If you do not know,
select "Don't know". Please, note that you are only being asked to answer what you
know. We are not asking you to find out any additional information. "Don't know"
should be used when you have no basis to give the answer. You may select "Don't know"
as often as necessary. [Note: original emphasis]

Items:
1. SECD was presented to staff as a serious, integrated commitment that will take years
to implement fully and to realize full behavioral and academic benefits (not as extra or
“feel-good”)
2. An SECD committee (either including a coordinator or lead person) exists and is active in sustaining SECD. That is, SECD does not rely solely on the existence of a single coordinator.

3. The coordinator/committee helps the staff troubleshoot issues that arise in DSACS implementation and supports the teaching staff by easing access to materials and strategies.

4. There exists a core group of individuals (beyond the school coordinator/committee) who strongly support SECD and are very involved in it.

5. There are systematic opportunities for staff (more than just coordinator, at least a team of people) to reflect upon the DSACS experience and future directions for SECD.

6. There is a clear message from the school administrator that staff are expected to implement and reinforce SECD.

7. The school administrator allows for specific planning time for staff to work on DSACS implementation and planning.

8. The school administrator is aware (by official or unofficial monitoring) of the implementation of DSACS plans in the building.

9. Some type of data tracking is utilized to indicate whether there are changes in student behaviors or not.

10. Quantifiable data on student behavior (e.g., # of office referrals, # of service hours completed) is collected in order to justify funding and time allotted to SECD.

11. Members of the SECD Team have been involved in continuing professional development around SECD issues since their initial training.
12. The school is moving towards (or has achieved) effective and reliable in-house staff training for SECD. That is, the school is not exclusively depending on "outsiders" to train current and new staff on how to deliver SECD

13. The school does not passively receive information from the DSACS Liaison. That is, the SECD Team initiates contact with the DSACS Liaison and/or looks up SECD information on their own (at least some of the time)

14. The school has shown originality in using SECD principles in a way that makes the principles "their own" and not merely a flat, “carbon copy”

15. The school has partnered with community agencies to help increase access to funds and/or opportunities for SECD
Appendix B

Coding System for the SITS

The following coding system will be used to determine the implementation scores for schools using the Schools Implementing Towards Sustainability (SITS) Scale. As it will probably be harder to score 100% at Time 2 than at Time 1, two alternative scoring systems will be computed for Time 2 to make the Time 2 scores equivalent to the Time 1 scores. For Time 2 Version B will be used for calculating the changes across time, but Version A will be used for all other analyses.

**Time 1 Coding:** “False” = 0, “True” = 1, “Don’t know” (DK) = 0. DK has to be coded as 0 in order to run reliability analyses. This does not affect the SITS-Tot score below, as a missing value is equivalent to a 0 value with how it is calculated.

**Time 2 Coding (version A):** “False, not at all true” = 0, “Somewhat true” = 1, “True” = 2, “Very true” = 3, DK = 0. This is the primary coding system for Time 2.

**Time 2 Coding (version B):** “False, not at all true” = 0, and all three types of true are coded as 1. This is the alternative coding system for Time 2.

**SITS-Tot** = Total implementation score. All the true values are added together. The minimum possible score is 0 and the maximum possible score is 15 for Time 1 and for the alternative coding of Time 2. The minimum possible score is 0 and the maximum possible score is 45 for the main coding of Time 2. Schools are penalized for DK responses as they are mathematically equivalent to “False” responses, which is
unfortunate but necessary. Conceptually, having one piece of the system of implementation does not guarantee that the other piece also exists as each piece represents a separate factor. Pragmatically, too many DKs were chosen by respondents to use a mean replacement method of calculation and still have a meaningful end product.
## Appendix C

Comparison of the CASEL Implementation Rubric and the SITS Scale

<table>
<thead>
<tr>
<th>CASEL Rubric Step/Factor*</th>
<th>Corresponding SITS Item*</th>
<th>Corresponding SITS Item</th>
<th>Corresponding SITS Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Principal commits to</td>
<td>1. serious, integrated,</td>
<td>6. admin says</td>
<td>8. admin aware</td>
</tr>
<tr>
<td>school-wide SEL</td>
<td>long-term</td>
<td>implement/reinforce</td>
<td>of impl</td>
</tr>
<tr>
<td>2. Engages stakeholders and form steering committee</td>
<td>2. active committee</td>
<td>15. partnered w/ comm</td>
<td></td>
</tr>
<tr>
<td>3. Develop and articulate shared vision</td>
<td>6. admin says</td>
<td>implement/reinforce</td>
<td></td>
</tr>
<tr>
<td>4. Conduct needs and resources assessment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Develop action plan</td>
<td>5. opp to reflect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Select evidence-based program</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Conduct initial staff development</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Launch SEL instructions in classrooms</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Expand instruction and integrate SEL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>school-wide</td>
<td>5. opp to reflect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Revisit activities and adjust for improvement</td>
<td>11. cont’ proff devel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A. Provide ongoing professional development</td>
<td>12. in-house staff training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B. Evaluate practices and outcomes for improvement</td>
<td>9. data track stu behave</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Develop infrastructure to support SEL</td>
<td>3. committee aids staff w/ probs/impl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D. Integrate SEL framework school-wide</td>
<td>4. core group outside committee planning time</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E. Nurture partnerships with families &amp; communities</td>
<td>5. opp to reflect</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F. Communicate w/ stakeholders</td>
<td>15. partnered w/ comm</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*See Appendix A for the complete wording of the SITS items and Ji et al. (2008) for the complete wording of the CASEL Implementation Rubric steps/factors. The numbers of
the CASEL Implementation Rubric represent the steps and the letters represent the ongoing sustainability factors (what this thesis calls the system of implementation). Steps 1 through 2 are the readiness phase (what this thesis calls the preparation phase), steps 3 through 6 are the planning phase (what this thesis calls the initial delivery phase), and steps 7 through 10 are the implementation phase (what this thesis calls the initial delivery phase and the continuing delivery phase).

Only two items on the SITS (i.e., Q13 and Q14) do not correspond well to a CASEL Implementation Rubric step/factor. One item (i.e., Q6) corresponds well to two CASEL items. One item (i.e., Q15) corresponds well to three CASEL items. One item (i.e., Q5) corresponds well to four CASEL items.

Five steps on the CASEL Implementation Rubric (i.e., Q4, Q6, Q7, Q8, and Q9) do not correspond well to a SITS item. However, these five steps are a normal part of the DSACS services provided (see Setting under METHOD for details) and their inclusion would have been redundant in this setting. For instance, the Climate Survey (see Climate Survey under Measures) accesses student and staff needs (step 4) and the 8-week cycle goals are specific action plans (step 5).
## Appendix D

Mapping the Phases and Levels onto the SITS Scale

<table>
<thead>
<tr>
<th>System of Implementation</th>
<th>Phase**</th>
<th>Level***</th>
</tr>
</thead>
<tbody>
<tr>
<td>(SITS items*)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. active committee</td>
<td>1. prep</td>
<td>2. school</td>
</tr>
<tr>
<td>1. serious, integrated, long-term</td>
<td>1. prep</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>6. admin says implement &amp; reinforce</td>
<td>1. prep</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>7. admin allows planning time</td>
<td>2. initial</td>
<td>2. school</td>
</tr>
<tr>
<td>4. core group outside committee</td>
<td>2. initial</td>
<td>2. school</td>
</tr>
<tr>
<td>3. committee aids staff w/ probs/impl</td>
<td>2. initial</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>8. admin aware of impl</td>
<td>2. initial</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>9. data track stu behave</td>
<td>2. initial</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>10. quantifiable data on stu behavior</td>
<td>2. initial</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>13. committee looks up info</td>
<td>2. initial</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>5. opp to reflect</td>
<td>3. con't</td>
<td>2. school</td>
</tr>
<tr>
<td>11. con’t prof devel</td>
<td>3. con't</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>12. in-house staff training</td>
<td>3. con't</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>14. orig in using principles</td>
<td>3. con't</td>
<td>2.5 school/district</td>
</tr>
<tr>
<td>15. partnered w/ comm.</td>
<td>3. con't</td>
<td>4. community</td>
</tr>
</tbody>
</table>

* Items: See Appendix A for the complete wording of the SITS items. Items in this appendix are sorted by phase and level rather than by their order in the survey.
** Phases: “1. prep” = Preparation Phase, “2. initial” = Initial Delivery Phase, “3. con't” = Continuing Delivery Phase

*** Levels: “2. school” = school ecological level, “2.5 school/district” = school and district ecological level, “4. community” = community ecological level

Note: While some items are labeled as pertaining to both the school and district ecological levels, they should only be considered to be tapping into the school level for the results presented in this thesis. Schools coordinators and liaisons were asked to answer the SITS items for their individual school. The label of school and district is given in this table because conceptually those items could be measurable at the district level. For instance, in-house training for SECD (Q12) could be organized and executed at the district level for all its schools rather than at each individual school throughout the district. Similarly, data could be tracked (Q9, Q10) at the school level but analyzed at the district level. And, leadership (Q1, Q6, Q8) can occur at both levels: the superintendent of the district can emphasize the importance of SECD implementation to the school principals and the school principals can emphasize its importance to their teachers.
Appendix E
Student Climate Survey

Answer Choices:
Participants chose “Disagree a Lot” [1], “Disagree a Little” [2], “Neither Agree Nor Disagree” [3], “Agree a Little” [4], or “Agree a Lot” [5] for each item. Items with an [R] at the end were reverse coded.

Instructions:
Please help us learn about your school. Your opinion will help us to try to make your school a better place.

Your name is not on the form, so no one at your school will know how you answered these questions. Please be as honest as you can.

There are no right or wrong answers. Your opinion is all that counts.

Please fill in all answers on your bubble sheet. Remember to fill in the bubble completely and stay inside the lines!

Thank you for your help!!!

Items:
1. Students treat classmates with respect.
2. Students exclude those who are different. [R]
3. Students help each other, even if they are not friends.
4. When students do something hurtful, they try to make up for it.
5. Students try to get other students to follow school rules.

6. Students work well together.

7. Students are disrespectful toward their teachers. [R]

8. Students help new students feel accepted.

9. Students pick on other students. [R]

10. Students are willing to forgive each other.

11. Students resolve conflicts without fighting, insults, or threats.

12. Students like being in this school.

13. Students are involved in helping to solve school problems.

14. Students can talk to their teachers about problems that are bothering them.

15. In this school, students don't feel like they learn anything useful. [R]

16. Teachers go out of their way to help students who need extra help.

17. Teachers in this school like to come here.

18. In this school you can count on adults to try to make sure students are safe.

19. Teachers are unfair in their treatment of students. [R]

20. Students here have a lot of school pride.
Appendix F

Staff Climate Survey

Answer Choices:
Participants chose “Strongly Disagree” [1], “Somewhat Disagree” [2], “Neither Agree Nor Disagree” [3], “Somewhat Agree” [4], or “Strongly Agree” [5] for each item. Items with an [R] at the end were reverse coded.

Instructions:
Our district is participating in the Developing Safe and Civil Schools project, designed to help improve academic performance, social-emotional growth, and non-violent conflict resolution in our students.

As part of this project, we are working with a team from Rutgers to learn about your school’s environment over time.

The purpose of this survey is to determine areas for improvement in your school, not to find fault with any particular department, area, or individual. Your answers are anonymous, so no one in this school or school district will know how you answered these questions. Please be as honest as you can.

Items:
1. In the last 3 months, how would you rate your overall satisfaction with your experience in this school?
2. Students treat classmates with respect.
3. Students exclude those who are different. [R]

4. Students help each other, even if they are not friends.

5. When students do something hurtful, they try to make up for it.

6. Students try to get other students to follow school rules.

7. Students work well together.

8. Students are disrespectful toward their teachers. [R]

9. Students help new students feel accepted.

10. Students pick on other students. [R]

11. Students are willing to forgive each other.

12. Students resolve conflicts without fighting, insults, or threats.

13. Students like being in this school.

14. Students are involved in helping to solve school problems.

15. Students can talk to their teachers about problems that are bothering them.

16. In this school, students don’t feel like they learn anything useful. [R]

17. Teachers go out of their way to help students who need extra help.

18. Teachers in this school like to come here.

19. You can count on adults to try to make sure students are safe.

20. Teachers are unfair in their treatment of students. [R]

21. Students here have a lot of school pride.

22. Students are disrespectful toward their parents in the school environment. [R]

23. Teachers treat parents with respect.

24. Parents show respect for teachers.
25. In their interactions with students, staff act in ways that demonstrate the character qualities the school is trying to teach.

26. Faculty and staff treat each other with respect.

27. This school shows appreciation for the efforts of faculty and staff.

28. The rules and sanctions relating to discipline in this school are well understood by both staff and students.

29. I receive support from my colleagues.

30. I am hesitant to approach the administration in this school to discuss concerns or grievances. [R]

31. There is a structure and ongoing process that provides feedback on my work performance.

32. Faculty and staff are involved in helping to make school decisions.

33. Teachers frequently discuss and share teaching methods and strategies with each other.

34. There is an agreed philosophy on discipline in this school.

35. Teachers are overloaded with work in this school. [R]

36. The school’s administrators don’t really know the problems faced by staff members. [R]

37. I am happy with the quality of feedback I receive on my work performance.

38. There is good communication between staff members in this school.

39. There is support from the administration in this school.

40. The morale in this school is high.
Table 1

<table>
<thead>
<tr>
<th>Surveys Counts</th>
<th>2008-2009</th>
<th>2009-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>**Active DSACS Schools * **</td>
<td>150</td>
<td>126</td>
</tr>
<tr>
<td>Schools w/ Liaison SITS Survey</td>
<td>138 (92%)</td>
<td>120 (95.2%)</td>
</tr>
<tr>
<td>Schools w/ Coordinator SITS Survey</td>
<td>71 (47.3%)</td>
<td>63 (50.0%)</td>
</tr>
<tr>
<td>Schools w/ Both SITS Surveys</td>
<td>69 (46.0%)</td>
<td>61 (48.4%)</td>
</tr>
<tr>
<td>Schools w/ Student Climate Survey **</td>
<td>84 (56.0%)</td>
<td>52 (41.3%)</td>
</tr>
<tr>
<td>Schools w/ Staff Climate Survey **</td>
<td>79 (52.7%)</td>
<td>34 (27.0%)</td>
</tr>
<tr>
<td>Schools w/ Both Climate Surveys **</td>
<td>69 (46.0%)</td>
<td>29 (23.0%)</td>
</tr>
<tr>
<td>Schools w/ Liaison SITS &amp; Student Climate Surveys ***</td>
<td>84 (56.0%)</td>
<td>52 (41.3%)</td>
</tr>
<tr>
<td>Schools w/ Liaison SITS &amp; Staff Climate Surveys ***</td>
<td>79 (52.7%)</td>
<td>34 (27.0%)</td>
</tr>
<tr>
<td>Schools w/ Coordinator SITS &amp; Student Climate Surveys ***</td>
<td>43 (28.7%)</td>
<td>29 (23.0%)</td>
</tr>
<tr>
<td>Schools w/ Coordinator SITS &amp; Staff Climate Surveys ***</td>
<td>38 (25.3%)</td>
<td>17 (13.5%)</td>
</tr>
</tbody>
</table>

* Between the 2008-2009 and the 2009-2010 academic year, 18 new schools became active in DSACS and 42 old schools became inactive. There are 157 unique schools with at least one SITS survey.

** Only counting climate surveys if at least one SITS survey exists for that school.

*** During the 2009-2010 academic year the State of New Jersey dramatically cut the budgets of all schools, which resulted in less resources (e.g., time and personnel) to complete climate surveys.
Table 2

*Climate Survey Demographics*

<table>
<thead>
<tr>
<th></th>
<th>Students in 2008-2009</th>
<th>Students in 2009-2010</th>
<th>Staff in 2008-2009</th>
<th>Staff in 2009-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Participants</td>
<td>30,114</td>
<td>19,973</td>
<td>3,997</td>
<td>2,037</td>
</tr>
<tr>
<td>Average Number of</td>
<td>386.1 (37-1683)</td>
<td>443.8 (66-1778)</td>
<td>50.6 (13-205)</td>
<td>56.6 (12-155)</td>
</tr>
<tr>
<td>Participants (Range:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>min-max)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Percent of</td>
<td>50.0% (39.5-61.0%)</td>
<td>49.5% (40.9-57.9%)</td>
<td>82.8% (57.6-100%)</td>
<td>82.6% (51.6-100%)</td>
</tr>
<tr>
<td>Female Participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Range: min-max)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Percent of</td>
<td>56.2% (12.1-100%)</td>
<td>54.5% (6.0-99.1%)</td>
<td>n/a*</td>
<td>n/a*</td>
</tr>
<tr>
<td>Non-White Participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Range: min-max)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average Percent of</td>
<td>n/a*</td>
<td>n/a*</td>
<td>69.9% (5.9-97.2%)</td>
<td>72.5% (23.7-93.5%)</td>
</tr>
<tr>
<td>Respondents who were</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Only Teachers</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Range: min-max)**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* n/a = not applicable. The data does not exist because it was not on the survey for that respondent type (i.e., It was only on the student or staff climate survey).
** If a staff participant holds a position as both a teacher and something else, that staff member is not counted for this row.
### Table 3

**School Demographics**

<table>
<thead>
<tr>
<th></th>
<th>2008-2009</th>
<th>2009-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tier 1</td>
<td>30</td>
<td>21</td>
</tr>
<tr>
<td>Tier 2</td>
<td>67</td>
<td>60</td>
</tr>
<tr>
<td>Tier 3</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>Cohort 1</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>Cohort 2</td>
<td>27</td>
<td>22</td>
</tr>
<tr>
<td>Cohort 3</td>
<td>58</td>
<td>50</td>
</tr>
<tr>
<td>Cohort 4</td>
<td>46</td>
<td>44</td>
</tr>
<tr>
<td>Primary Schools</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Elementary Schools</td>
<td>51</td>
<td>44</td>
</tr>
<tr>
<td>Elementary/Middle Schools</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>Middle Schools</td>
<td>33</td>
<td>34</td>
</tr>
<tr>
<td>Middle/High Schools</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>High Schools</td>
<td>23</td>
<td>19</td>
</tr>
<tr>
<td>Public Schools</td>
<td>128</td>
<td>115</td>
</tr>
<tr>
<td>Low District Factor Grouping (DFG) (indication of Low SES)</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td>Average Enrollment in 2008-2009** (Range: min-max)</td>
<td>700.5 (52-2033)</td>
<td>729.6 (154-2414)</td>
</tr>
</tbody>
</table>

* Primary = no grades above 3; Elementary = mostly grades 1-5; Elementary/Middle = mix of grades 1-8; Middle = mostly grades 6-8; Middle/High = mix of grades 6-12; High = mostly grades 9-12.

** Enrollment data for 2009-2010 is unavailable. However, the total enrollment of students does not change much between years.
Table 4

*Number of Schools that Increased and Decreased their SITS scores across Time*

<table>
<thead>
<tr>
<th></th>
<th>Decrease (%)</th>
<th>Increase (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RCI (year)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DSACS Liaisons (n = 101)</td>
<td>3 (3.0%)</td>
<td>7 (6.9%)</td>
</tr>
<tr>
<td>RCI (.80)</td>
<td>5 (5.0%)</td>
<td>11 (10.9%)</td>
</tr>
<tr>
<td>Half SD</td>
<td>13 (12.9%)</td>
<td>39 (38.6%)</td>
</tr>
<tr>
<td><strong>School Coordinators (n = 34)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RCI (year)</td>
<td>0 (0%)</td>
<td>3 (8.8%)</td>
</tr>
<tr>
<td>RCI (.80)</td>
<td>0 (0%)</td>
<td>4 (11.8%)</td>
</tr>
<tr>
<td>Half SD</td>
<td>0 (0%)</td>
<td>17 (50.0%)</td>
</tr>
</tbody>
</table>

Note: Changes are computed using the alternative coding system for Time 2. RCI (year) = Reliable Change Index assuming that the test-retest reliability for a normal length is equal to the 6-12-month-long test-ret; RCI (.80) = Reliable Change Index assuming that the test-retest reliability for a normal length is equal to .80; Half SD = SITS score changed by half of a standard deviation or more.
Figure 1

Effects of Implementation Assistance by Experts on Implementation & Sustainability Scores