Illinois Horizon Schools: Initial Research Findings from the Illinois Schools to Watch Program

Steven B. Mertens
Illinois State University, School of Teaching & Learning, Normal, IL, USA, smertens@ilstu.edu

Kristina N. Falbe
Illinois State University, School of Teaching & Learning, Normal, IL, USA

Abstract

Currently, there is a paucity of research concerning the National Forum to Accelerate Middle Grades Reform’s Schools to Watch national school improvement program. Of the handful of existing studies, very few have examined or addressed student learning or achievement. In this study we systematically collected and analyzed data from the 34 Illinois Horizon Schools to Watch (IHSTW) schools participating in the National Forum’s Schools to Watch (STW) program. Using multiple sources of data, we examined the demographic characteristics and teacher-reported levels of best middle grades school practices in the IHSTW middle level schools that have participated in the STW program since 2003. In addition, we conducted an analysis of state-level, standardized student achievement test scores for participating IHSTW schools. We found that IHSTW schools that have participated in the program for longer periods of time demonstrated higher levels of middle school practices and standardized student achievement outcomes.

Keywords: middle level schools, school improvement, Schools to Watch, Self-Study Rating Rubric, student achievement

Established in 1997, the National Forum to Accelerate Middle Grades Reform (the Forum) is an alliance of educators, researchers, national associations, and officers of professional organizations and foundations committed to promoting academic performance and healthy development of young adolescents (National Forum to Accelerate Middle-Grades Reform, 2019). To accomplish its goal of improved academic and developmental outcomes for all students in the middle grades, the Forum identifies and disseminates best practices, articulates and promotes effective policies, recognizes and develops enlightened leadership, and informs and engages the public (National Forum to Accelerate Middle-Grades Reform, 2019).
2019). In 2001, the National Forum launched its Schools to Watch® (STW) program designed to identify and recognize schools that were on a positive trajectory of school improvement. As part of the STW program, the Forum developed a set of criteria for identifying high-performing middle level schools, created tools to help schools use the criteria for school improvement, expanded the program to 17 states, and recognized nearly 500 STW schools across the country.

The National Forum’s criteria for identifying high-performing middle level schools is derived from its vision statement: “In order to prepare students to be lifelong learners ready for college, career, and citizenship, the National Forum seeks to make every middle-grades school academically excellent, responsive to the developmental needs and interests of young adolescents, and socially equitable” (National Forum to Accelerate Middle-Grades Reform, 2019, para 2). In addition to the criteria related to academic excellence, developmental responsiveness, and social equity, the Forum added an additional criterion focused on school organizational structures and processes.

In 2003, Illinois became one of three inaugural states to be selected to implement the STW state program. The STW program in Illinois is called Illinois Horizon Schools to Watch (IHSTW), and it is facilitated and operated by the Association of Illinois Middle-Grades Schools (AIMS), a statewide network of middle grade schools initiated in 1990. Since 2003, AIMS has worked with nearly 60 middle level schools in Illinois interested in the IHSTW program; currently there are 34 middle-grade schools active in the IHSTW program.

Similar to the implementation of the STW program in other states, Illinois schools interested in participating in the IHSTW program are required to submit an initial designation application. In addition to basic demographic information about the school (e.g., location, enrollment, demographics, grade levels, percent free/reduced lunch), the application asks schools to describe how they address each of the STW criteria (i.e., academic excellence, developmental responsiveness, social equity, and organizational structures and processes). As part of the application process, faculty and administrators in the school are required to complete an online survey focused on the STW rubric and its components. The online Schools to Watch® Self-Study and Rating Rubric serves as a type of needs assessment for the school, addressing needs or gaps between current practices, procedures, and policies and desired best practices, procedures, and policies. The results of the STW Self-Study and Rating Rubric are returned to the school to inform the official application process.

The National Forum determined that an initial designation as a School to Watch is only valid for three years. Prior to the end of the third year, schools are required to apply to be re-designated and continue as a School to Watch. The re-designation process consists of completing a re-designation application and the School to Watch® Self-Study and Rating Rubric. A unique aspect of the STW program is its focus on continuous school improvement efforts characterized by a trajectory toward success. It is an intentional focus on continuous improvement practices and provides a systemic process for engaging all faculty and administrators in self-reflection and collective discussions about improvement.

**Purpose**

The purpose of our study was to systematically examine the various data collected as part of the IHSTW program to determine if the participating schools are a representative sample of Illinois middle level schools and whether the program is effective in increasing and sustaining levels of best practices and standardized student achievement test scores. We collected, organized, and analyzed the data related to the IHSTW program including program applications, the STW Self-Study and Rating Rubric data, and school-level, standardized student achievement data. We sought to answer the following four research questions:

RQ1: Are the demographic characteristics of the IHSTW schools comparable to the demographic characteristics of all other middle level schools in Illinois?

RQ2: Are the school improvement practices assessed on the STW Self-Study and Rating Rubric higher for IHSTW schools that have been re-designated two or more times than for schools that have been designated or re-designated once or twice?

RQ3: Are the standardized achievement test scores for students in IHSTW schools higher compared to scores for students in non-IHSTW middle level schools in Illinois?

RQ4: Is there any difference in mean student standardized achievement test scores based on designation status?

**Literature Review**

The middle school concept is an educational approach aligned with the unique developmental needs of young adolescents (10–15 years old). Advocacy organizations such as the Association for Middle Level Education...
(AMLE), the Carnegie Council on Adolescent Development (CCAD), and the National Forum have worked to define what effective middle level practices are and to delineate best practices and tenets for schools designed for young adolescent learners (Carnegie Council on Adolescent Development [CCAD], 1989; Jackson & Davis, 2000; National Forum to Accelerate Middle-Grades Reform, 2019; National Middle School Association [NMSA], 2010). Although these organizations do not espouse the exact same set of characteristics, they all support practices predicated on the need for school to be responsive to the developmental needs of young adolescent learners (Falbe, 2014).

These beliefs focus upon young adolescent developmental needs, academic challenges, school culture and climate, and organizational practices and can be identified in key middle level practices like interdisciplinary teaming, advisory, common planning time, interdisciplinary and integrated curriculum, and authentic and exploratory curriculum (CCAD, 1989; Jackson & Davis, 2000; National Forum to Accelerate Middle-Grades Reform, 2019; NMSA, 2010). Developed from a vision of the National Forum, the STW program seeks to identify and define school excellence at the middle level. Lipsitz and West (2006) explained that the National Forum holds a central belief that “young adolescents are capable of learning and achieving at high levels”; they are “dedicated to improving schools for middle-grades students across the country” (p. 57).

To date, very little empirical research has been conducted on the STW schools or the STW program, and only one study has used the national sample of STW schools (Mertens & Flowers, 2016). Extant research on STW has typically focused on a state-specific program or school(s) or has used a subset of the STW national sample to address a particular issue or topic, like academic or achievement outcomes or social equity and developmental responsiveness (see Cook & Faulkner, 2010; Cook et al., 2009; Falbe, 2014, 2015; McEwin & Greene, 2011; Mertens & Flowers, 2016; Parke et al., 2017).

In one of the larger studies, McEwin and Greene (2011) sought to “provide a longitudinal perspective on the degree of implementation of key middle grades programs and practices” (p. 49). Their 2009 study included a national random sample of 827 public middle level schools and a sample of 101 highly successful middle level schools (HSMS). These HSMS included both Schools to Watch and the National Association of Secondary School Principals’ Breakthrough Middle Schools. Administrators in the schools were asked to complete an electronic survey to provide data about their schools and to express their opinions on selected middle level topics. Compared to their random sample of national middle level schools, McEwin and Greene (2010, 2011) found that the principals in the HSMS sample reported more frequent use of interdisciplinary team organization, flexible block scheduling, exploratory-type courses, and cooperative and inquiry-based learning. In addition, the HSMS were more likely to have a higher number of common planning periods per week, advisory programs, and higher percentages of core teachers with a specific middle level teacher certification.

In another large-scale, state-level study, Cook et al. (2009) examined a stratified random sample of 40 Kentucky middle level schools and 10 Kentucky Schools to Watch schools to determine if there were any differences in perceived level of implementation of the middle school concept and student academic achievement. Data was collected via an electronic survey from 568 certified personnel in the 50 schools, 55% from STW schools and 45% from non-STW schools. The authors found that respondents in STW schools reported a slightly higher perceived level of implementation of tenets of the middle school concept—teacher preparation and certification, advisory, teaming, and common planning time—as well as higher levels of standardized student achievement (Cook et al., 2009).

Conceptual Framework

This study examines the IHSTW schools through the lens of the National Forum’s vision, specifically its STW criteria, and the best practices it has developed as part of its STW Self-Study and Rating Rubric (National Forum to Accelerate Middle-Grades Reform, 2019). Although this vision and rubric are essential to the work of the National Forum, it is a framework rooted in and influenced by the middle school concept as delineated in the following seminal publications: Turning Points: Preparing American Youth for the 21st Century (CCAD, 1989); Turning Points 2000: Educating Adolescents in the 21st Century (Jackson & Davis, 2000), and This We Believe: Keys to Educating Young Adolescents (NMSA, 2010).

The recommendations contained in these seminal publications, while representing separate organizations and entities, are quite similar with each
representing a conceptual framework for school improvement and/or reform. In developing its vision, the National Forum to Accelerate Middle-Grades Reform built upon and was influenced by these national recommendations for middle level reform. The Forum drafted a vision statement that could be used to define and identify high-performing middle level schools and is comprised of four interrelated tenets: academic excellence, social equity, developmental responsiveness, supported by organizational structures and processes. After developing their vision statement, the Forum recognized the need to develop a set of criteria for schools to use to evaluate and assess their progress toward school improvement. The Forum “concluded that in order to pursue these priorities, high-performing schools must be learning organizations that establish norms, structures, and organizational arrangements that will support and sustain their trajectory toward excellence” (Lipsitz & West, 2006, p. 58). The 37 criteria developed for the four tenets are complementary and interdependent and have been used by the Forum for nearly 20 years to identify high-performing schools, schools that sit “at the intersection of academic excellence, developmental responsiveness, and social equity” (Lipsitz & West, 2006, p. 58).

The framework for this study is based on the seminal publications noted above and the National Forum’s vision and corresponding STW criteria. We investigated the levels of middle level best practices in IHSTW schools compared to statewide middle level schools, and we examined student outcomes in the form of standardized student achievement data of IHSTW schools and the state. The results of our study will help to validate the use of the Schools to Watch criteria for middle level school improvement. Answers to these research questions will also provide evidence that the STW designation identifies schools that can serve as models of excellence for other schools, thus refuting the claims that the middle school concept is ineffective (e.g., Yecke, 2005). Our study also builds on prior work that examined and measured (Alexander, 1968; Alexander & McEwin, 1989; Cook et al., 2009; Lee & Smith, 1993; McEwin et al., 1996, 2003; McEwin & Greene, 2010).

The middle school concept, as outlined in the seminal publications noted above and in the National Forum’s vision and criteria, maintains that young adolescents deserve an education that is appropriate to their developmental needs and provides social equity and academic excellence, and that such an education can be realized through application of the middle school concept (CCAD, 2000; Jackson & Davis, 2000; National Forum to Accelerate Middle-Grades Reform, 2019; NMSA, 2010). That understanding and position is what leads us to use the STW framework as a measure of middle level practices.

**Methodology**

We used a non-experimental research design and included secondary data sources gathered from three primary sources, (a) the IHSTW applications (initial and re-designation), (b) the STW Self-Study and Rating Rubric data, and (c) school-level, standardized student achievement data in the form of the Illinois Report Card data files.

**Study Population**

The study population was comprised of the 34 IHSTW schools that participated in the IHSTW program through the 2018–19 school year. As noted earlier, the IHSTW designation status is valid for three years, after which schools are required to apply for re-designation status. Since the inception of the IHSTW program in 2003, 34 Illinois middle level schools have received an initial designation:

- 24 schools have been re-designated once (minimum of 3 years in the program);
- 16 schools been re-designated twice (6 years minimum);
- 13 schools been re-designated three times (9 years minimum); and
- 4 schools have been re-designated four times (minimum of 12 years).

**Data Sources**

This study used secondary data sources in the form of the IHSTW program applications, online survey data in the form of the STW Self-Study and Rating Rubric (collected as part of the IHSTW application/re-designation process), and school-level, standardized student achievement data from the Illinois Report Card data files (available for download from Illinois State Board of Education website).

An additional data source was the 2016–17 National Center for Education Statistics’ Core of Common Data (CCD) files available for download from the U.S. Department of Education’s NCES website (National Center for Education Statistics [NCES], n.d.). At the time of the analyses, the
most current and available CCD were from the 2016–17 year. The CCD files contain basic demographic information for all K–12 public schools in the U.S. including location, enrollment, grade configurations, student ethnicity, free/reduced lunch percentages, and other demographic characteristics. These data were used to triangulate with the IHSTW application data to confirm various school demographic information. The school demographic information enabled us to determine the representativeness of the current sample of IHSTW schools compared to all middle level schools in the state.

Data from the online STW Self-Study and Rating Rubric were obtained, by request, from the Center for Prevention Research and Development (CPRD) at the University of Illinois, which manages and coordinates this annual data collection on behalf of the National Forum. CPRD requires that within each school, 80% of classroom teachers complete the online survey for the schools to receive feedback and for their data to be included in any subsequent analyses. The online STW Self-Study and Rating Rubric survey provides anonymous, individual teacher and administrator data about the four constructs of the National Forum’s vision: academic excellence (25 questions), social equity (27 questions), developmental responsiveness (31 questions), and organizational structures and processes (24 questions). The individual STW rubric items for each of these four constructs are combined into four composite scores by calculating a mean score for each of the constructs. The data from the online STW Self-Study and Rating Rubric enabled us to address our third research question by determining if IHSTW schools that have been re-designated multiple times have higher mean composite scores than the schools that have been designated or re-designated once or twice. To date, 34 IHSTW schools have participated in the online STW Self-Study and were subsequently designate or re-designated as an IHSTW. The dataset for the current study included 1,893 cases (teachers and administrators).

The third source of data for the study was school-level, standardized student achievement data in the 2018 Illinois State Board of Education Report Card (Illinois State Board of Education [ISBE], n.d.). In 2015, Illinois adopted the national Partnership for Assessment of Readiness for College and Careers (PARCC) standardized achievement test. Due to criticisms of the PARCC assessment, Illinois adopted a new state assessment, the Illinois Assessment of Readiness (IAR), in 2019 (Thayer, 2019). The IAR incorporates the Common Core standards and is administered in English language arts and mathematics. The report cards include school/building information such as student enrollments, instructional setting, finances, and academic performance/assessment data.

The Schools to Watch® Self-Study and Rating Rubric Construct
The STW Self-Study and Rating Rubric was developed by the National Forum in 2006 and has been updated, revised, and refined over the course of the past decade. The anonymous online rubric is completed by teachers and school administrators as part of the STW school application process. As noted earlier, the majority (80%) of classroom teachers within a school are required to complete the survey for their data to be considered valid. Some of the survey questions address demographic information (e.g., role in school, grades taught, subjects taught). The rubric was designed to solicit ratings from teachers and administrators across the four STW constructs: (a) academic excellence, (b) developmental responsiveness, (c) social equity, and (d) organizational structures and processes. Within each construct there are approximately 8 to 10 criteria with examples. Teachers rate each of the 37 STW criteria on a scale from 1 to 4:

1. 1 – Poor quality, low level of implementation, new program;
2. 2 – Fair quality, mixed implementation, immature practice, sporadic by some;
3. 3 – Good quality, maturing but not fully implemented by all teachers; or
4. 4 – High quality, complete, mature, and coherent implementation.

The individual STW rubric items for each of the four constructs are combined into four composite scores by calculating a mean score for each construct. To measure the reliability of the STW Self-Study and Rating Rubric, specifically the internal consistency of the items used to create the composite scores, Cronbach alphas were calculated from prior studies and ranged from 0.96 to 0.98 (Center for Prevention Research and Development [CPRD], University of Illinois, 2015, 2018).

CPRD at the University of Illinois, who partnered with the National Forum to coordinate the online data collection, has conducted evaluations for federal
grants received by the National Forum. Utilizing quasi-experimental research designs, CPRD’s evaluation reports have examined the implementation of the STW constructs (academic excellence, developmental responsiveness, social equity, and organizational structures and processes) through the use of the STW Self-Study and Rating Rubric (CPRD, 2015, 2018). These evaluation studies have utilized the online rubric data to address levels of implementation of specific characteristics or attributes of the STW criteria and are limited to state-specific or sub-samples of the STW schools. This study is the first to utilize multiple sources of STW data (i.e., applications, online rubric data, and standardized student achievement data) to examine the impact of the STW program and criteria on the state-wide population of IHSTW schools.

Data Analysis
The analyses in the current study consisted of descriptive statistics, means analysis, and t-tests or analyses of variance (ANOVA), depending on the number of categories for the independent variables. Descriptive analyses of the school demographic characteristics (school location, grade configuration, enrollment, student ethnicity/race, and percentage of free/reduced-price lunch students) were conducted of the IHSTW schools and compared to the statewide population of middle level schools to determine the representativeness of the IHSTW schools (see Tables 1–4). This analysis enabled us to address our first research question: Are the demographic characteristics of the IHSTW schools comparable to the demographic characteristics of all other middle level schools in Illinois?

To address our second research question, we had to conduct a means analysis to determine if the variances within the range of scores for each STW construct (academic excellence, developmental responsiveness, social equity, and organizational structures and processes) is acceptable or if there are outlier scores that will need further investigation. Subsequently, we conducted statistical tests (t-test or ANOVA) to test for statistically significant differences in the group mean scores between newly designated IHSTW schools compared to IHSTW schools that have been re-designated multiple times. In these analyses, designation status (initial designation, first re-designation, second re-designation, etc.) served as the independent variable and the mean scores of the STW constructs served as the dependent variables. Through these analyses we were able to assess if longevity in the program (and continued implementation to the STW constructs) demonstrated statistically significant gains in the mean scores for the STW constructs.

We addressed the third and fourth research questions by conducting analyses of the school-level, standardized student achievement data in the form of the Illinois School Report Card data. Through an analysis of the Illinois Assessment of Readiness test scores—the percentage of students “meeting” or “exceeding” expectations—we compared school-level, state achievement test scores in English language arts and mathematics of the current IHSTW schools and the population of Illinois middle level schools. We conducted statistical tests of the group mean standardized student achievement scores (t-tests or ANOVA, depending on the categories of the independent variable). In these analyses, school type (IHSTW school or state school) and designation status served as the independent variables and the mean, school-level achievement scores served as the dependent variables.

Findings
Our report of the findings follows the three data sources: (a) the school demographic data, (b) the STW Self-Study and Rating Rubric survey data, and (c) the school-level, standardized student achievement data.

Table 1
Comparison of Enrollment and Free/Reduced Lunch Status: Illinois Middle Grade Schools Vs. IHSTW Schools

|                      | State (Min) | State (Max) | State (Mean) | IHSTW (Min) | IHSTW (Max) | IHSTW (Mean) |
|----------------------|-------------|-------------|--------------|-------------|-------------|--------------|
| School Enrollment    | 21          | 1478        | 413.7        | 194         | 1478        | 735.4        |
| F/R Lunch Status     | 0.0%        | 100%        | 51.7%        | 2.6%        | 99.7%       | 42.8%        |
Table 2
Comparison of Student Race: Illinois Middle Grade Schools Vs. IHSTW Schools

| Race                | State (Max) | State (Mean) | IHSTW (Max) | IHSTW (Mean) |
|---------------------|-------------|--------------|-------------|--------------|
| % White             | 100.0       | 54.3         | 95.9        | 54.9         |
| % Black             | 100.0       | 17.9         | 90.3        | 11.0         |
| % Hispanic          | 99.5        | 20.3         | 99.0        | 23.8         |
| % Asian             | 85.2        | 3.7          | 36.3        | 7.2          |
| % Pacific Islander  | 5.0         | 0.1          | 0.3         | 0.03         |
| % Native American   | 48.9        | 0.3          | 0.6         | 0.2          |
| % Multiracial       | 39.6        | 3.4          | 10.4        | 2.9          |

Table 3
Comparison of Building Grade Configurations: Illinois Middle Grade Schools Vs. IHSTW Schools

| Grade Configs | State (%) | IHSTW (%) |
|---------------|-----------|-----------|
| PK/K-8        | 17.3      | 16.1      |
| 5–8           | 2.6       | 3.2       |
| 6–8           | 10.3      | 80.6      |
| Other         | 69.8      |           |

Table 4
Comparison of School Location: Illinois Middle Grade Schools Vs. IHSTW Schools

| School Location | State (%) | IHSTW (%) |
|-----------------|-----------|-----------|
| City            | 23.9      | 32.3      |
| Suburban        | 42.2      | 41.9      |
| Town            | 13.5      | 9.7       |
| Rural           | 20.4      | 16.2      |

School Demographic Variables
Our analyses of the demographic variables derived from the IHSTW applications and the data from the CCD found that the IHSTW schools were quite similar in their demographic characteristics to all other Illinois middle level schools. Compared to the statewide mean school enrollments, IHSTW schools had a higher average student enrollment ($M = 735.4$) compared to the statewide average student enrollment ($M = 413.7$) and a significantly higher minimum number of students ($N = 194$) for IHSTW schools versus 21 for the state (see Table 1). The IHSTW schools had higher, on average, student enrollments than all other state middle grades schools. This is important to note as the IHSTW program is not selectively choosing middle grades schools with low student enrollments.

A 2016 quasi-national study of a sample of STW schools found that the demographics of the schools were quite varied and diverse (Mertens & Flowers, 2016). Using a longitudinal dataset of 166 STW schools in 15 states, it was found that the free/reduced-price (F/R) lunch percentage ranged from 0–96% with designated STW schools having 44% F/R lunch students and non-designated schools having 46%. In the current study, the average (mean) number of F/R lunch students in IHSTW schools was 43% compared to 52% for the state, a nine percent difference (see Table 1). Although not identical in F/R lunch percentages, the IHSTW schools are not significantly lower (43% vs. 52%).

In examining the racial composition of students in the IHSTW schools compared to the statewide schools, we found the percentages of students to be very comparable (see Table 2). IHSTW schools had a slightly lower percentage of Black students compared to the state (11% vs. 18%). However, compared to the statewide schools, IHSTW schools had a slightly higher percentage of Hispanic students (24% vs. 20%) and Asian students (7% vs. 4%), and the percentage of White students was nearly identical (55% vs. 54%).

When examining grade configuration, it is important to note that schools participating in the STW program
(nationally and within Illinois) are not required to be solely comprised of middle grades configurations (e.g., 5–8, 6–8, 7–8). The STW program is designed to focus on the middle grade levels regardless of the school/building grade configurations. Currently there are a number of national STW schools with varying grade configurations including, for example, K–8, 4–8, and 6–12 schools. Within Illinois, there are currently over 40 variations of school/building grade configurations containing middle grade levels (i.e., sixth, seventh, eighth). As the IHSTW program focus on the middle grades, we compared the range of grade configurations in the IHSTW schools and the statewide schools. Currently the IHSTW schools contain one of three grade configurations, 5–8, 6–8, and P/K–8. As can be seen in Table 3, there are comparable percentages of grade configurations for the IHSTW and state schools, 17% have PK/K–8 grade configurations and 3% have 5–8 grade configurations. The overwhelming majority (81%) of the IHSTW schools have a 6–8 grade configuration compared to only 10% for the state. This is not surprising as the IHSTW program is designed to address school reform at the middle grades level; therefore, we would expect the majority of the participating IHSTW schools to have a 6–8 grade configuration. In comparison, slightly more than 10% of Illinois schools have a 6–8 grade configuration.

Lastly, we wanted to examine school location within the state. School location classifications (i.e., city, suburban, town, rural) were determined from the CCD files which use the U.S. Census Bureau classifications. Table 4 contains the comparison of the IHSTW school locations to the statewide schools. The percentage of suburban schools is nearly identical (42%) and there is higher percentage of IHSTW schools in city locations compared to the state (32% vs. 24%). However, compared to the state, the percentage of IHSTW schools is lower in towns (10% vs. 14%) and rural areas (16% vs. 20%). The majority (42%) of the IHSTW schools are located in Cook County and 74% of the IHSTW schools are in the northern region of the state.

**STW Self-Study and Rating Rubric Survey**

As mentioned earlier, the STW Self-Study and Rating Rubric is an online survey completed by teachers and administrators prior to being designated or re-designated as a STW school. When completing the survey, teachers are asked to rate their own classroom practices on a 4-point scale across the four STW rubric components: academic excellence (25 items), social equity (27 items), developmental responsiveness (31 items), and organizational structures and processes (24 items). Of the 1,893 school personnel who completed the 2018–19 online survey, 82% were classroom teachers; the remaining 18% were administrators, counselors, media specialists, and curriculum coordinators/instructional coaches. Eighty-eight percent of teachers taught in sixth, seventh, or eighth grade. For the purposes of the subsequent analyses, only respondents who identified themselves as classroom teachers were selected, thus reducing the overall sample to 1,322 individuals.

Subscale averages were calculated for each of the four STW rubric constructs. Each school in the dataset was assigned a designation status based on the number of times the school had been re-designated as part of the IHSTW review process. An analysis of variance (ANOVA) was conducted for each of the four STW rubric constructs with the subscale average scores (dependent variable) and designation status (independent variable). The results of all four of the ANOVAs were found to be statistically significant at \( p < .001 \) (see Table 5). Post hoc analyses, using the Bonferroni correction procedure, found that the primary differences were between schools receiving initial designation vs. schools that had been re-designated for multiple years. The ANOVA results suggest that the longer schools are involved in the IHSTW program (i.e., receiving multiple re-designations), the more likely teachers are to rate their levels of practice at a higher level. Effect sizes, in the form of eta squared (\( \eta^2 \)), were calculated for each of the four STW rubric subscales; a small to medium effect was found for the four constructs.

**School-Level, Standardized Student Achievement Data**

As noted above the standardized student achievement data for this study were derived from the 2018 Illinois School Report Card data file from the Illinois State Board of Education (IS, n.d.); student achievement is reported as the percentage of students meeting or exceeding expectations. For comparison purposes, we selected only Illinois schools which contained middle grade levels (e.g., K–8, K–12, 5–8, 6–8). There were 42 grade configuration categories containing a seventh grade and one contiguous grade level resulting in a sample of 1,712 schools (excluding the 34 IHSTW schools).

In order to address our third research question (Are standardized achievement test scores for students in IHSTW schools higher compared to students in non-IHSTW middle level schools in Illinois?), we conducted an independent-samples \( t \)-test to evaluate
the difference in the group standardized student achievement mean scores for IHSTW schools and all other Illinois middle grade schools in sixth, seventh, and eighth grade English language arts and mathematics standardized student achievement scores. As can be seen from Table 6, the differences in mean scores for all groups were statistically significant at the $p < .05$ level or better (all but one were significant at $p < .002$). Cohen’s $d$ effect size values were calculated; sixth grade language arts had a medium effect size and all other effect sizes were large (Table 6).
Table 6
**Group Differences in Illinois Achievement (PARCC) Test Scores by School Type (IHSTW Vs. State)**

| Achievement Tests: Grade & Content | IHSTWM (SD) | StateM (SD) | t-value (df) | p-value | Cohen’s d |
|-----------------------------------|------------|------------|-------------|---------|----------|
| 6th Grade English Language Arts   | 39.9 (17.7)| 30.8 (21.5)| −2.46 (1567)| 0.014   | 0.46     |
| 7th Grade English Language Arts   | 47.1 (21.1)| 35.2 (22.3)| −3.08 (1399)| 0.002   | 0.55     |
| 8th Grade English Language Arts   | 44.7 (18.6)| 31.7 (21.2)| −3.54 (1397)| 0.000   | 0.65     |
| 6th Grade Mathematics             | 34.0 (19.0)| 23.4 (19.0)| −3.22 (1567)| 0.001   | 0.56     |
| 7th Grade Mathematics             | 38.3 (19.0)| 25.5 (19.4)| −3.80 (1399)| 0.000   | 0.67     |
| 8th Grade Mathematics             | 40.0 (16.9)| 25.4 (18.8)| −4.48 (1397)| 0.000   | 0.82     |

To address our fourth research question, we conducted an ANOVA comparing the mean achievement scores by IHSTW designation status. There were no statistically significant differences in mean achievement scores by designation status. However, there were some interesting observations. For English language arts, initial designation schools generally had the lowest mean achievement scores. Schools re-designated for a fourth time (minimum 15 years in program) had the highest mean achievement scores for sixth (50.9%) and eighth grade (57.0%) (see Table 7). Schools re-designated twice had the highest mean ELA scores for seventh grade (61.8%). Similar trends were observed for mathematics scores; initial designation schools had the lowest mean achievement scores and schools re-designated for a fourth time had the highest mean achievement scores for all grade levels (43.2% in sixth grade, 47.0% in seventh grade, and 48.9% in eighth grade).

**Limitations and Delimitations**

There were some limitations to the current study. First, the number of Illinois middle level schools participating in the IHSTW program is not particularly large, only 34 schools through the 2019–20 academic year. According to the Illinois State Board of Education’s school report card data, there were 1,712 schools in Illinois with middle level grades of which the 34 IHSTW schools represent approximately two percent. The lack of any notable statistical significance when comparing standardized student achievement scores by designation status could be attributed to the smaller sample of 34 IHSTW schools. The sample size could also be considered a delimitation as schools are required to apply to participate in the STW program. Despite earnest efforts on an annual basis to recruit new schools, only a handful of schools typically apply to the IHSTW program each year. Another limitation is the self-reported nature of the of the STW Self-Study and Rating Rubric survey. The online survey is voluntarily completed by teachers, administrators, and staff. We believe the ratings from the teacher survey were reliable as they were anecdotally corroborated by independent IHSTW reviewers after visiting the schools and conducting observations and interviews.

**Discussion**

Despite concerns that the STW schools may not be demographically representative of middle level schools nationally, prior research utilizing a quasi-national study of STW schools found that this not be the case (Mertens & Flowers, 2016). In this study, we found that IHSTW schools have a slightly lower percentage of free/reduced lunch students as compared to all other middle grade schools in the state (43% vs. 52%). We would suggest that this nine percent difference is minimal and does not discount the representativeness of the IHSTW schools to the statewide population of middle grades schools. With a few exceptions, we believe the school demographics between the IHSTW schools and the statewide middle grades schools are quite comparable. Although the IHSTW schools may not yet serve as an exact representative sample of the state’s middle school population, we believe it is similar enough to warrant valid comparisons.
Table 7
ANOVA Results for Illinois Achievement Scores by Designation Status

| Rubric Criteria/Construct | Mean | SD  | F (df)      | p     | $\eta^2$ |
|---------------------------|------|-----|-------------|-------|---------|
| **6th Grade English Language Arts** |      |     |             |       |         |
| Designated (10 schools)   | 36.3 | 16.4| 1.02 (4,29) | 0.415 | 1.55    |
| 1st re-designation (6 schools) | 36.0 | 19.5|            |       |         |
| 2nd re-designation (5 schools) | 49.5 |  5.6|            |       |         |
| 3rd re-designation (9 schools) | 36.5 | 22.7|            |       |         |
| 4th re-designation (4 schools) | 50.9 | 13.2|            |       |         |
| **6th Grade Mathematics** |      |     |             |       |         |
| Designated                | 28.4 | 16.7| 0.49 (4,29) | 0.745 | 1.35    |
| 1st re-designation        | 37.1 | 19.7|            |       |         |
| 2nd re-designation        | 36.6 | 10.2|            |       |         |
| 3rd re-designation        | 32.8 | 27.1|            |       |         |
| 4th re-designation        | 43.2 | 12.2|            |       |         |
| **7th Grade ELA**         |      |     |             |       |         |
| Designated                | 37.1 | 16.7| 1.22 (4,29) | 0.322 | 1.77    |
| 1st re-designation        | 48.0 | 25.8|            |       |         |
| 2nd re-designation        | 61.8 | 12.5|            |       |         |
| 3rd re-designation        | 47.9 | 24.5|            |       |         |
| 4th re-designation        | 50.3 | 20.7|            |       |         |
| **7th Grade Mathematics** |      |     |             |       |         |
| Designated                | 30.8 | 14.6| 0.91 (4,29) | 0.469 | 1.55    |
| 1st re-designation        | 43.3 | 22.1|            |       |         |
| 2nd re-designation        | 45.4 |  6.2|            |       |         |
| 3rd re-designation        | 35.6 | 26.3|            |       |         |
| 4th re-designation        | 47.0 | 12.6|            |       |         |
| **8th Grade ELA**         |      |     |             |       |         |
| Designated                | 36.4 | 16.4| 1.57 (4,29) | 0.208 | 1.87    |
| 1st re-designation        | 45.2 | 19.0|            |       |         |
| 2nd re-designation        | 56.1 | 17.3|            |       |         |
| 3rd re-designation        | 41.9 | 21.2|            |       |         |
| 4th re-designation        | 57.0 | 10.6|            |       |         |
(Continued)
The analyses of the \textit{STW Self-Study and Rating Rubric} survey data presented some very interesting results. With one exception, the results of the ANOVAs clearly suggest that the longer schools are involved with the IHSTW program the higher the mean levels of all four of the teacher self-reported STW rubric constructs: academic excellence, developmental responsiveness, social equity, and organizational structures and processes. We should note the one anomaly that exits in this analysis. In general, the mean scores for each of the four constructs increase by designation status with the exception of schools that have been re-designated for third time. These schools, representing the second largest group of schools in the study (nine schools), had a slight decrease in their mean scores across all four of the STW constructs. Although their mean scores are slightly lower than schools re-designated a second time, their standard deviations all fall within the range of standard deviations for each STW construct. At this time, we cannot offer any plausible reason why this is the case; however, this phenomenon requires investigation in a future study. Lastly, schools designated for the first time have the lowest levels on the STW rubric constructs. With the exception of the schools redesignated a third time, the longer schools were engaged in the IHSTW program, the higher their levels of implementation of the STW criteria (see Table 5).

The analyses of the school-level standardized student achievement data provided some very positive results. We were interested in comparing the school-level achievement data for IHSTW schools to the statewide population of middle grade schools. The results of the independent-samples $t$-test found that the differences in mean scores for all groups were statistically significant. This suggests that students in IHSTW schools outperformed the statewide population of middle grades school in both English language arts and mathematics across all grade levels, sixth, seventh, and eighth. This is a remarkable, but not surprising, finding. As noted above, the results of the \textit{STW Self-Study and Rating Rubric} survey found levels of best practices, as measured by the STW criteria were, with the exception of the nine schools re-designated for a third time, consistently higher the longer a school has participated in the IHSTW program. Thus, it should not come as a surprise that students in these IHSTW schools outperformed students in other statewide middle grade schools.

Our other research question dealing with the standardized student achievement test data pertained specifically to the IHSTW schools. We were interested in determining if there were any statistically significant differences in school-level, achievement data based on the designation status of the IHSTW schools. In other words, does length of time participating in the IHSTW program improve school-level achievement scores. The ANOVA did not produce any statistically significant results, suggesting that length of time in the program does not have any impact on achievement. However, the relatively small sample size of 34 schools may also be a contributing factor in the lack of any statistical significance in the ANOVA results. We did observe that schools just starting in the program had the lowest mean achievement scores, for both English language arts and mathematics. In addition, and with only one exception, we found that the schools re-designated for a fourth time had the highest mean achievement scores for all grade levels (the exception was seventh grade ELA). These are promising results
and will be the focus of future study as the sample of IHSTW schools hopefully increases.

Conclusions
In conclusion, we believe that the results of this study are very promising and demonstrate the effectiveness of the IHSTW program. Our analyses of the outcome data—teacher reports of best practices as represented by the four STW constructs and the student standardized student achievement data—clearly indicate that the longer schools participate and engage in the IHSTW program, the more improved their classroom practices were as assessed by the STW Self-Study and Rating Rubric. The STW program is designed for schools to focus on continual, long-term school improvement efforts. The IHSTW schools participating in the program for numerous years clearly show improvements in teacher ratings of academic excellence, developmental responsiveness, social equity, school organizational structures and processes, and student achievement. It is important to reiterate that students in IHSTW schools outperformed students in all other middle level schools in the state in both English language arts and mathematics’ standardized student achievement tests.

In addition, although there were only 34 IHSTW schools in Illinois through 2019–20, it is astounding to note that 16 schools have been involved in the IHSTW program for over six continuous years, 14 schools for over nine years, and four schools for over 12 years. It is also quite likely that over the course of their time in the program, many of these IHSTW schools saw a turnover in their building principal and teaching staff. However, through a shared common vision—the STW vision—these schools were able to persist in their trajectory of school improvement.

The results from this study are comparable to other large-scale studies that examined various components of the middle school concept combined with student outcomes as assessed through standardized achievement test data (e.g., Cook et al., 2009; Falbe, 2014, 2015; McEwin & Greene, 2011). Similar studies have relied on measures of the middle school concept developed by the study’s authors and administered to middle school principals or teachers (e.g., Cook et al., 2009; McEwin et al., 1996, 2003; McEwin & Greene, 2011). Our study is different in that it used self-reported teacher data collected from the STW Self-Study and Rating Rubric, specifically designed to reliably assess the 37 criteria within the four tenets of the STW vision.

The current study focused on a group of 34 Illinois middle grade schools participating in the statewide IHSTW program. Based on the results of this study we see the IHSTW program as a viable means of sustainable school improvement and would recommend that more schools become involved in the IHSTW program. We would also recommend that schools embrace the tenets of the STW vision and seek to implement the STW criteria into their school practices, programs, and policies, including professional development. We would encourage state-level and local policy makers to support the IHSTW program and seek ways to expand it into middle level schools across all regions of the state. Lastly, we strongly recommend continued research with STW schools. Similar state-level studies could be initiated in the hopes of eventually conducting a quasi-national study of STW schools.

The results of this study clearly demonstrate that high-performing middle level schools can successfully address the STW tenets of academic excellence, developmental responsiveness, and social equity and provide supportive learning environments for young adolescents. In recent years, the national STW program has garnered some attention from legislators across the county. Now that an empirical basis demonstrating the effectiveness of the IHSTW program is building, perhaps the results will influence statewide policies and funding for middle level schools. We hope the results of this study support the work of the dedicated teachers and administrators currently working with young adolescents in middle level schools in Illinois and across the country. We believe our study’s findings clearly demonstrate that continuous, positive school change is possible for Illinois middle level schools through their participation in the IHSTW program. The program is a clear example of how high-performing middle level schools can provide academically excellent, developmentally responsive, and socially equitable educational learning experiences for all young adolescents.

Acknowledgments
We acknowledge and thank the following people and organizations for providing access the Illinois Horizon Schools to Watch data: John Viano, Executive Director, Association for Illinois Middle-Grade Schools (AIMS) and the AIMS
board, and Ericka Uskali, former Executive Director of the National Forum to Accelerate Middle Grades Reform, for providing access to the IHSTW data and accompanying documentation and their permission to use the data in this research effort. A special thanks to Nancy Flowers, Director of Research Programs at the Center for Prevention Research and Development, University of Illinois, for providing us with electronic copies of the STW Self-Study and Rating Rubric data for the IHSTW schools and for her feedback and suggestions on an earlier draft of this article.

Funding

This work was funded, in part, through a University Research Grant from Illinois State University.

ORCID

Steven B. Mertens http://orcid.org/0000-0002-5518-0235
Kristina N. Falbe http://orcid.org/0000-0002-2456-7947

References

Alexander, W. M. (1968). A survey of organizational patterns of reorganized middle schools. United States Department of Health, Education, and Welfare. https://files.eric.ed.gov/fulltext/ED024121.pdf

Alexander, W. M., & McEwin, C. K. (1989). Schools in the middle: Status and progress. National Middle School Association.

Carnegie Council on Adolescent Development. (1989). Turning points: Preparing American youth for the 21st century. Carnegie Corporation.

Center for Prevention Research and Development, University of Illinois. (2015). Schools to Watch: School Transformation Network. A U.S. Department of Education Investing in Innovation (i3) Development Grant: Final evaluation report. University of Illinois.

Center for Prevention Research and Development, University of Illinois. (2018). Middle-Grades Leadership Development (MLD) Project: A U.S. Department of Education Investing in Innovation (i3) Development Grant: Final evaluation report. University of Illinois.

Cook, C. M., & Faulkner, S. A. (2010). The use of common planning time: A case study of two Kentucky Schools to Watch? Research in Middle Level Education Online, 34(2), 1–12. https://doi.org/10.1080/19404476.2010.11462075

Cook, C. M., Faulkner, S. A., & Kinne, L. J. (2009). Indicators of middle school implementation: How do Kentucky’s Schools to Watch measure up? Research in Middle Level Education Online, 32(6), 1–10. https://doi.org/10.1080/19404476.2009.11462059

Falbe, K. N. (2014). The relationship between Schools to Watch designation and academic achievement: A study of Colorado, New York, Ohio, and Virginia [Unpublished doctoral dissertation]. Mercer University.

Falbe, K. N. (2015, April). The relationship between middle schools with Schools to Watch designation and academic achievement [Paper presentation]. American Educational Research Association, Chicago, IL, USA.

Illinois State Board of Education. (n.d.). Illinois report card. Illinois State Board of Education https://www.isbe.net/reportcard

Jackson, A. W., & Davis, G. A. (2000). Turning points 2000: Educating adolescents in the 21st century. Teachers College Press.

Lee, V., & Smith, J. (1993). Effects of school restructuring on the achievement and engagement of middle-grades students. Sociology of Education, 66(3), 164–187. https://doi.org/10.2307/2112735

Lipsitz, J., & West, T. (2006). What makes a good school? Identifying excellent middle schools. Phi Delta Kappan, 88(1), 57–66. https://doi.org/10.1177/003172170608800110

McEwin, C. K., Dickinson, T., & Jenkins, D. (2003). America’s middle schools in the new century: Status and progress. National Middle School Association.

McEwin, C. K., Dickinson, T. S., & Jenkins, D. M. (1996). America’s middle schools: Practices and progress—a 25-year perspective. National Middle School Association.

McEwin, C. K., & Greene, M. (2011). The status of programs and practices in America’s middle schools: Results from two national studies. Association for Middle Level Education.

McEwin, C. K., & Greene, M. W. (2010). Results and recommendations from the 2009 national surveys of randomly selected and highly successful middle level schools. Middle School Journal, 42 (1), 49–63. https://doi.org/10.1080/00940771.2010.11461750

Mertens, S. B., & Flowers, N. (2016, April). The National Schools to Watch program: Characteristics and implementation levels of successful middle level schools [Paper presentation]. American Educational Research Association, Washington, DC, USA.

National Center for Education Statistics. (n.d.). Common core of data: America’s public schools.
Institute of Education Sciences. https://nces.ed.gov/ccd/

National Forum to Accelerate Middle-Grades Reform. (2019). About the Forum. National Forum to Accelerate Middle-Grades Reform. https://www.middlegradesforum.org/about-us

National Middle School Association. (2010). This we believe: Keys to educating young adolescents. National Middle School Association.

Parke, C. S., Generett, G. G., & Ramos, R. A. (2017). Describing characteristics of Pennsylvania’s “Schools to Watch:” Focusing on social equity and developmental responsiveness. Middle School Journal, 48(4), 13–26. https://doi.org/10.1080/00940771.2017.1343056

Thayer, K. (2019, March 6). Illinois scraps controversial PARCC test in favor of shorter exam with new name. Chicago Tribune. https://www.chicagotribune.com/lifestyles/ct-life-iar-parcc-state-testing-20190227-story.html

Yecke, C. P. (2005). Mayhem in the middle: How middle schools have failed America—and how to make them work. Thomas B. Fordham Institute.