## *Effectiveness of Pearson's SuccessMaker Mathematics for Students with Disabilities*

**Steven K. McKissick**
**Texas A&M University**

### *Abstract*

SuccessMaker mathematics is an instructional learning system rooted in behaviorist instructional theory. Previous research efforts have left much to be desired and have produced inconsistent results. Recent research for this program appears to be tapering off, despite advances in technology signaling integration of concepts from other theoretical positions. A quasi-experimental review of data from a sample of students ($N$ = 1186) from a central Texas school district over a five-year period was reviewed. Multivariate analysis of variance identified that changes in state testing performance were not linked to program use. Changes in the rate of academic achievement were found to exist between usage groups. Students who met or exceeded usage recommendations (>20 hours of use) were found to have significantly greater rates of achievement (ES: $d$ = 1.02). Recommendations for further studies and limitations of the current study are provided.

## *Effectiveness of Pearson's SuccessMaker Mathematics for Students with Disabilities*

Educators and researchers have spent more than thirty years investigating a class of technological interventions known as instructional learning systems (ILS). An ILS has been described as a "software program that provides tutorial instruction at several grade levels and keeps extensive records of student progress on networked computer systems" (Kulik, 2002, p. 1). Bailey (1992) expanded this description by identifying five key characteristics that separate an ILS from other instructional technology: (a) ability to target specific instructional objectives and connect these to specific lessons; (b) potential for integration into other curricula; (c) span multiple grade levels, possibly in multiple content areas; (d) the use of a networked computers; and (e) collection of student performance records. Though the National Council of Teachers of Mathematics (2000) has emphasized the inclusion of instructional technology in classrooms, the implementation and use of an ILS is more involved than the use of calculators or interactive smartboards. Various ILS technologies have been reviewed to include products developed by Wicat Systems and Jostens Learning Corporation, as well as programs such as Plato, Prescription Learning, and SuccessMaker (Becker, 1992).

Because ILS use is frequently treated as a supplemental curriculum, recommendations for ILS use have not always been followed. A number of ILS programs come with recommendations for minimum student usage (Gee, 2008; Manning, 2004). Failure to integrate the ILS with existing classroom curriculum has resulted in ILS usage of about 15-30% of program recommendations (van Dusen & Worthen, 1995). A matrix to evaluate technology implementations contrasted this "unacceptable use" with "ideal use" wherein the ILS is used "as a tool for regularly accomplishing classroom instructional objectives" (Mills & Ragan, 2000, p. 28). Because of such variation, Slavin (1987) urged that time spent using the program be a factor in determining the effectiveness of an ILS.

SuccessMaker is an ILS for which a historical review may be necessary to identify relevant research. The program is rooted in the work of Suppes and Zancotti at Stanford University in the late 1960s (Kulik, 1994; Wood, 2004). Out of their work came the Computer Curriculum Corporation (CCC) and, ultimately, this program. The company was purchased by Simon & Schuster in 1990 (Manning, 2004). Pearson acquired Simon & Schuster and its holdings, including SuccessMaker, in 1998 (Pearson Digital Learning, n.d.). Previous research with the program has identified it as Stanford-CCC, SuccessMaker, SuccessMaker Enterprise, or even by a portion of the product such as Math Concepts and Skills (Manning, 2004). Given the changes in ownership and name, it is doubted that all previous relevant studies were identified in previous ILS meta-analyses.

A discussion about the nature of SuccessMaker Mathematics (SMM) is helpful for identifying an underlying theoretical framework. Students begin their use of SMM with an initial placement assessment designed to identify grade level skills. This process may take up to three hours (Pearson Digital Learning, 2012) or approximately 300 questions (Wood, 2004). Students may begin this initial placement at either their enrolled grade level or a level determined by the teacher managing the student's use of the program. Students are presented with questions that increase or decrease in difficulty depending on the accuracy of student responses. A branching algorithm is used to work through various skill strands and grade levels (Svoboda, Jones, van Vulpen, & Harrington, 2012). Students may work on skills at their ability level, in 15 strands of content (Pearson Digital Learning, 2004), with difficulty contingent on student success. Additionally, teachers may assign specific skill units to students instead of having students work only on grade-level skills. SMM, as anticipated by Bailey's (1992) description of an ILS, maintains an ongoing record of student skill capabilities and program usage, allowing the teacher to produce up-to-date records of student use and progress when needed. SMM also incorporates a regular review of previously mastered skills into student work to ensure continued understanding (Wood, 2004).

SMM is an interactive program within a multimedia environment. Students are provided with audio and video material regarding a particular concept or skill. Students have access to virtual tools such as a highlighter and sticky notes to keep students active during learning (Pearson Education, 2013). No research studies were found that examined these particular tools for effectiveness. SMM provides immediate feedback for student responses. A "cognitive coach who offers hints and insights" (p. 6) is provided when a student answers incorrectly. This use of a multimedia environment for learning has been found to improve student comprehension during instruction (Bransford, Sherwood, Hasselbring, Kinzer, & Williams, 1990).

**Theoretical Considerations**

The behaviorist definition of learning is the acquisition of a new behavior. A person learns what is practiced, and learning prepares the student to demonstrate "specific responses to particular stimuli rather than general responses to vague stimuli" (Schiro, 2013, p. 63). The learner is considered an active participant in the learning process, and exhibition of learned behaviors is necessary for continued learning (Ormrod, 2014). Shaping occurs as increasingly complex or difficult behaviors are presented to the learner. Schiro (2013) noted that even the most complex tasks are considered by behaviorists as compositions of discrete simpler skills that can be taught. Immediate feedback is necessary, and technology increases that immediacy. Learning is self-paced; not all learners will acquire the same skill at the same speed or in the same number of discrete trials.

SMM has its foundations in behaviorism through programmed instruction. Programmed instruction, as developed by Skinner (1986), is a specific application of behaviorist principles built on the early work of Thorndike and Pressey. Material to be learned should be presented in small increments to reduce the likelihood of error. Material is arranged by complexity, and learners enter at the highest level at which they can demonstrate mastery (Svoboda et al., 2012). The learner is presented with a question in response to some stimulus, and the teacher (or, for SMM, the program) awaits a response. The student is provided differential feedback based on the response. Failure to respond correctly in SMM may result in continued exposure to the same skill with additional support from the "cognitive coach" or a change in skill or skill level following multiple failures suggesting frustration. Students experiencing consistent success may experience an increase in the grade level of skills presented through a process known as branching (Joyce, Weil, & Calhoun, 2009). The present level of student ability is identified as the skill level where the student's performance plateaus, and instruction is provided at that level.

Programmed instruction has changed significantly as technology has changed. The rise and fall in favor with programmed instruction has been directly linked to these technological changes (Svoboda et al., 2012). In early years, programmed instruction led to an over-reliance on technology which, coupled with a limited range of stimulating media, led to student boredom (McDonald, Yanchar, & Osguthorpe, 2005). Rigid application of the principles of programmed instruction identified above has relaxed in later years (McDonald et al., 2005), and later programs and versions have been more interactive and student-directed (Cruthirds & Hanna, 1997). Current iterations of SMM have retained core principles of programmed instruction – success-driven increases in complexity, immediate feedback, and active participation – while sprinkling in tools more consistent with cognitive and constructivist frameworks.

Programmed instruction works, though research findings are inconsistent. Early meta-analytic research found that programmed instruction yielded an effect size of just over $d = .20$ (Kulik, Kulik, & Cohen, 1980), at the low end of Cohen's (1988) bracket for a small effect. Two years later, another meta-analysis determined that programmed instruction was no better than traditional instruction (Kulik, Schwalb, & Kulik, 1982), with an effect size for mathematics of $d = -.01$. Another early estimate of the effectiveness of computer-aided instruction, to include systems utilizing programmed instruction, yielded an effect size of $d = .57$ (Schmidt, Weinstein, Niemiec, & Walberg, 1985). Ormrod (2014) contends that programmed instruction remains viable for students with little previous success, including students with learning or behavior difficulties, as well as those for whom previous attempts at teaching and learning have proven unsuccessful. Behaviorist principles are well-established, though their application may be time-intensive and less than enjoyable.

Behaviorist strategies have demonstrated success with learning-disabled students (Zafiropoulou & Karmba-Schina, 2005). The reason may be attributed to the ability of computer-based interventions, such as SMM, to provide immediate feedback (Burton, Moore, & Magliare, 2008). Cooley (2007) proposed that students with mathematics disabilities be provided with step-by-step modeling of solving problems, frequent monitoring of progress, and the use of work sessions that are more frequent but less intense. Drill-and-practice models have been recommended (Pellegrino & Goldman, 1987) as a step towards building automaticity of skills (Cummings & Elkins, 1999). “Those who lack automaticity at the basic skills level exhaust their cognitive resources trying to recall math facts and, therefore, have few resources left for solving problems” (Wendling & Mather, 2009, p. 173). SuccessMaker Mathematics incorporates these recommendations and behaviorist principles, and it is anticipated that its use with students with learning and behavior disabilities should prove effective in increasing achievement levels.

Constructivist principles may also be seen in more recent iterations of SMM. By providing incremental increases in skills under review, SMM incorporates a mechanical version of Vygotsky’s (1978) zone of proximal development. According to Vygotsky, students learn best when challenged with skills at or slightly above their current ability level. By reinforcing previously learned skills, SMM also provides instructional scaffolds on an individual basis. Though the interpersonal contact and communication are absent from a true sociocultural position, communication via the cognitive coach and program use facilitated by the teacher may serve as surrogates. The communication provided by SMM during its instruction is a version of math dialogue akin to Richards’ (1996) “school math” characterized by rigidity and computational focus. This style is further characterized by an invitation-reply-discourse sequence; SMM provides a prompt-response-feedback communication loop. Mills and Ragan (2000) noted that the teacher should not be supplanted by any coaching provided through the ILS, and their ideal use of the ILS includes the teacher as an ongoing participant in the teaching process.

This author assumes a pragmatist position (Creswell, 2011; Creswell & Plano Clark, 2011) that avoids the discontinuities between the various theoretical frameworks above. Instead, pragmatism takes a “what works” approach and considers the question asked as more important than the underlying theory (Creswell, 2011; Creswell & Plano Clark, 2011; Tashakkori & Teddlie, 2003). This leads to a philosophical pluralism that allows for the inclusion of both behaviorist understandings of learning as well as constructivist epistemologies. Practicality, a focus on the outcomes and consequences of choices, is most valued (Cherryholmes, 1992; Tashakkori & Teddlie, 2003). The question being asked here is whether or not SuccessMaker is effective for improving mathematical learning for students with disabilities, not by what means it may do so.

**Previous Research Findings**

Though the research on instructional learning systems is rich, a historical review of SMM was more difficult. Possibly due to the variety of names by which the product has been called over the years, few primary source documents were found. Many studies that were identified had not been submitted to peer review through the journal publication process. A review of existing meta-analyses and research syntheses was undertaken. These studies are presented in Table 1, including selected details and effect sizes.

The studies presented in Table 1 are not without concern. Only six of the studies in Table 1 (Crawford, 1970; Delon, 1970; Mendelsohn, 1972; Ragosta, 1983; Suppes & Morningstar, 1969; Underwood, Cavendish, Dowling, Fogelman, & Lawson, 1996) have been subject to peer review. This increases the possibility that design flaws and inaccurate reporting may have led to erroneous results. Slavin and Lake (2008) identified design flaws in eleven studies, including Kirk (2003) and Underwood et al. (1996) presented here. A frequent design issue cited by Slavin and Lake (2008) was the lack of an adequate control group, though inadequate outcome measures and group equivalence were also noted as concerns among their excluded studies. Table 1 includes four institutional reports, and the most recent report included (Gatti, 2009) should be interpreted with caution as it appears to be research sponsored by the vendor for SMM.

Table 1
*Previous SuccessMaker Research*

| Study | Type of Publication | Location | Grade | Number of Subjects | Effect Size (*d*) |
|---|---|---|---|---|---|
| †Cranford (1976) | Dissertation | Mississippi | 5th – 6th | | .64 |
| †Crawford (1970) | Journal Article | California | 7th | 2 classrooms, 36 students | .10 |
| †Davies (1972) | Dissertation | California | 3rd – 6th | 240 students | .34 |
| †Delon (1970) | Journal Article | Mississippi | 1st | 5 classrooms, 99 students | 1.08 |
| Gatti (2009) | Institutional Report | 4 states (AZ, FL, MA, NJ) | 3rd, 5th | 8 schools, 792 students | .14 (for 3rd) .50 (for 5th) |
| Gee (2008) | Dissertation | Georgia | 3rd – 5th | 1 school, 180 students | .61 |
| *Hotard & Cortez (1983) | Institutional Report | Louisiana | 3rd – 6th | 2 schools, 190 students | .39 |
| †Jamison, Fletcher, Suppes, & Atkinson (1976) | Book Chapter | Mississippi | 1st – 6th | 12 schools, 600 students | .40 |
| Kirk (2003) | Dissertation | Tennessee | 2nd – 5th | 4 schools, 348 students | .84 (.93 for 5th) |
| Laub (1995) | Dissertation | Pennsylvania | 4th-5th | 2 schools, 314 students | .56 |

| Study | Type of Publication | Location | Grade | Number of Subjects | Effect Size ($d$) |
|---|---|---|---|---|---|
| Manning (2004) | Dissertation | Florida | $6^{th}$ | 1 school, 64 students | .75 |
| Manuel (1987) | Dissertation | Nebraska | $3^{rd}$-$6^{th}$ | 3 schools, 165 students | .06 |
| †Mendelsohn (1972) | Journal Article | New York | $2^{nd}$ – $6^{th}$ | 20 schools, 3,282 students | .49 |
| †Miller (1984) | Dissertation | Oregon | $5^{th}$ – $8^{th}$ | 15 schools, 577 students | .38 |
| Mintz (2000) | Dissertation | Alabama | $4^{th}$ – $5^{th}$ | 8 schools, 487 students | -.06 |
| †Palmer (1973) | Institutional Report | California | $4^{th}$ – $6^{th}$ | 3 schools, 171 students | .36 |
| †Prince (1969) | Institutional Report | Mississippi | $1^{st}$ – $6^{th}$ | 12 schools, 544 students | .64 |
| *Ragosta (1983) | Journal Article | California | $1^{st}$ – $6^{th}$ | 4 schools | .77 |
| †Suppes & Morningstar (1969) | Journal Article | California | $1^{st}$ – $6^{th}$ | 7 schools, 1896 students | .28 |
| Underwood, Cavendish, Dowling, Fogelman, & Lawson (1996) | Journal Article | United Kingdom | primary & secondary | 9 schools, 173 students | .40 |
| †Vincent (1977) | Dissertation | Ohio | $9^{th}$ – $12^{th}$ | 2 schools, 35 students | .34 |

Notes:_†Included in Kulik (1994) meta-analysis. *Included in Slavin and Lake (2008).

The lack of recent research regarding SMM is of concern. No peer-reviewed research was found that was been conducted in the past twenty years. The most recent research studies located were conducted by doctoral students as part of their dissertations (Gee, 2008; Kirk, 2003; Manning, 2004; Mintz, 2000). Though the research has investigated the same program, that program has doubtlessly changed over time to leverage new technological capabilities. At present, Pearson (2015) is advertising SuccessMaker 8 as the newest version of their software. It is unclear if differences between this version and previous versions are cosmetic, functional, or instructional. Given the ages of the studies listed in Table 1, it is reasonable to assume that the underlying theoretical framework relied heavily on programmed instruction (Svoboda et al., 2012).

An average effect size was found for the studies provided in Table 1, though certain assumptions were required. It was assumed that the sample in Gatti (2009) was equally split into two groups. The low effect size for Kirk (2003) was used as representative of her study given the concerns presented by Slavin and Lake (2008). The simple mean effect size found for studies in Table 1 was *d* = .46 (95%CI [.34, .57]). Using Cohen's (1988) suggestions regarding the interpretation of effect sizes, this result would be considered small. Removal of two significant outliers (Delon, 1970; Mintz, 2000) yielded a similar though slightly lower simple mean effect size of *d* = .41 (95%CI [.32, .50]). Notably, three of the highest effect sizes from these studies were from studies conducted in Mississippi nearly forty years ago (Cranford, 1976; Delon, 1970; Prince, 1969). Restricting this process to only studies conducted since 2000 did not result in significantly different results.

An additional evaluation of SMM research was conducted by Becker (1992). Results from 11 studies conducted during the 1980s were included, though citations for these studies were omitted by the author. As a consequence, locating Becker's original sources is unlikely. Becker's (1992) studies are described in Table 2. Becker included both sample sizes and effect sizes for the included studies, and a weighted mean effect size can be calculated. It is assumed that the sample size from the Calvert Co., Maryland study was equal for all three groups. The weighted mean effect size was *d* = .30 (95%CI [.12, .47]). This small effect size was statistically significant. However, the New York study contained nearly one-third of the cumulative sample in Becker's presentation, and the effect size for that study was a statistical outlier. Removal of this study and recalculation of the weighted mean effect size yielded an effect size of *d* = .45 (95%CI [.28, .63]). Studies done most recently generated effect sizes greater than the confidence interval for the revised mean effect size, suggesting a time-based effect perhaps tied to technology innovations.

Table 2
*Studies included in Becker (1992) Meta-Analysis*

| Study | Design | Location | Grade | Number of Subjects | Effect Size (*d*) |
|---|---|---|---|---|---|
| 1988-89 | Individual Change vs. Test Norms | Ft. Worth, TX | $1^{st} – 7^{th}$ | 120 students, ~25 hours use | 1.60 |
| 1988-89 | Individual Change vs. Test Norms | Omaha, NE | $2^{nd} – 6^{th}$ | 170 students, ~20 hours use | 1.30 |
| 1987-88 | Individual Change vs. Test Norms | Milwaukee, WI | $2^{nd} – 9^{th}$ | 600 students, ~40 hours use | .80 |
| 1987-88 | Individual Change vs. Test Norms | Aiken Co., SC | $2^{nd} – 8^{th}$ | 600 students, ~30 hours use | .70 |
| 1983-88 | Cohort Change to | Calvert Co., MD | $3^{rd}$, $5^{th}$, $8^{th}$ | 1,500 students, ~35 hours use | .10 ($3^{rd}$) .25 ($5^{th}$) |

| | | | | | |
|---|---|---|---|---|---|
| | Statewide Change | | | | .50 (8th) |
| 1983-86 | Individual Change vs. Test Norms | Calvert CO., MD | 4th – 6th | 653 students | .35 |
| 1977-80 | Random Assignment | Los Angeles, CA | 1st – 6th | 750 students, ~50 hours use | .26 |
| 1980-81 | Random Assignment | Lafayette Parish, LA | 3rd – 6th | 94 students, ~25 hours use | .19 |
| 1981-82 | Comparison Group | Portland, OR | 5th – 8th | 80 students, ~25 hours use | .30 |
| 1984-86 | Comparison Group | Rochester, NY | 4th – 6th | 2,600 students, 19 schools | .00 |
| 1984-85 | Comparison Group | Atlanta, GA | Elementary, Middle | 700 students, 7 schools ~25 hours use | .40 |

*Note*. Becker (1992) failed to provide authors for any of the studies included in his meta-analysis. Consequently, these studies are only descriptions of studies rather than identifications of studies. Most sample sizes are approximate.

A number of studies have been identified by previous authors but rejected for various reasons. Table 3 provides an overview of these studies. Many of the studies were rejected by Slavin and Lake (2008) for various reasons, though Pearson (2002) provided a collection of summaries for these. All of the studies in Pearson (2002) failed to provide sufficient statistical information from which to derive effect size information. Instead, percentiles and percentage passing rates appeared more frequently. None of the original studies could be found, though most appeared to be reports produced by either Pearson (vendor for SMM) or the school districts in which the product was used. None were submitted for peer review, and the likelihood of corporate authorship casts doubts as to the replicability of the studies. None of the studies were conducted in the past ten years.

Table 3
*Documents Not Included in Meta-Analytic Comparisons*

| Study | Type of Publication | Location | Grade | Number of Subjects | Data Provided |
|---|---|---|---|---|---|
| Crenshaw (1982) | Dissertation | | | | (a) |
| Donnelly (2004) | Presentation | | | | (b) |
| Humphries (1997) | Institutional Report | North Carolina | 3rd – 8th | 11 classrooms | percentiles |
| Laub & Wildasin (1998) | Institutional Report | Pennsylvania | 2nd – 6th | 6 schools, 522 students | percentiles, grade equivalents (a) |
| McWhirt, Mentavlos, Rose-Baele, & Donnelly, (2003) | Institutional Report | | | | (a) |
| Office of Research, Loudoun Co. Public Schools (1998) | Institutional Report | Virginia | 3rd – 5th | 3 schools, 254 students | qualitative overview |
| Phillips (2001) | Dissertation | | | | (c) |
| Simon & Tingey (2001) | Institutional Report | Florida | 4th – 5th | 12 schools, 459 students | FCAT results |
| Tingey & Simon (2001) | Institutional Report | California | 4th – 5th | 9 schools, 597 students | mean gains, normal curve equivalents (a) |
| Tingey & Thrall (2000) | Institutional Report | Florida | 4th – 5th | 12 schools | percentage comparisons (a) |
| Tuscher (1998) | Institutional Report | Pennsylvania | 3rd – 5th | 4 schools | SAT-9 percentiles (a) |
| Wildasin (1984) | Institutional Report | | | | (a) |

*Note*. All deficiency comments from Slavin & Lake (2008).
(a) Lack of an adequate control group. (b) Insufficient control group matching. (c) Inadequate outcome measure.

Previous research has suggested that SMM produces a small but significant effect on student achievement. Findings were inconsistent across types of studies (journal article vs. dissertation, etc.) as noted above. Study location may have even impacted findings. Research efforts regarding SMM may be tapering off; the last peer-reviewed article was published twenty years ago. Previous research has also focused on elementary mathematics performance. Only eight studies included students in $7^{th}$ or $8^{th}$ grades (traditional junior high or middle school grades). It is telling that the What Works Clearinghouse provides no judgment of the evidence-based effectiveness of SMM. More research is needed to determine if SMM truly yields an effect on students' mathematics achievement.

**Purpose of This Study**

National standards have been set through No Child Left Behind and Race to the Top by which schools are expected to demonstrate adequate yearly progress in mathematics. Students with disabilities have historically underperformed on these assessments relative to their non-disabled peers. As the number of students with disabilities grows, it becomes increasingly important to provide adequate supports for these students in order to meet state and national standards (Manning, 2004). Students with disabilities generally only make small achievement gains, especially during the middle school years (Graham, Bellert, Thomas, & Pegg, 2007). Pressures for students with disabilities, especially learning disabilities, to succeed are increasing (Martindale, Pearson, Curda, & Pilcher, 2005) while the gap between high and low achievers grows wider every year (Cawley, Parmar, Yan, & Miller, 1998).

Despite the research base for SMM outlined above, limited research exists to support its effectiveness for students with disabilities (Wood, 2004). Vockell and Mihail (1993) suggested that consistent computer-based instruction may provide students with disabilities a greater chance of success through development of automaticity and overlearning of concepts. It has also been suggested that technology should be integrated into mathematics instruction for all at-risk learners (Li & Edmonds, 2005). The aim of this study is to determine if SMM effectively improves mathematics achievement for students with disabilities.

## *Methods*

SuccessMaker Mathematics was purchased by a central Texas school district at the beginning of the 2010-2011 school year by the Special Education department. Consequently, schools were instructed that only students eligible for special educations services were to use the program. Licenses were purchased and given to all 12 middle schools in the district. Identification of specific students and development of a campus implementation plan was left to the campuses. Vendor recommendations to the district regarding yearly usage totals suggested that 20-25 hours of use per student should produce measurable achievement gains. Those recommendations are consistent with those currently provided by vendor representatives (D. Wayland, personal communication, January 28, 2016). A matrix of time usage estimates based on IP level and expected gain provided by the vendor (Pearson Education, 2012) was not available to the district at the start of their implementation. The array considers homogeneous clusters of students grouped by their IP level. Based on desired gain levels, usage levels are provided at three incremental levels of student success. The publication reads, in part, "Achieving the time in the $50^{th}$ percentile column will result in approximately one-half of students reaching at least that gain; achieving the time in the 75th percentile will result in approximately three-fourths of students reaching at least that gain" (Pearson Education, 2012). Given the wide range of achievement levels for students using SuccessMaker both district-wide and at each campus, the matrix was condensed to a yearly usage recommendation of approximately 20-25 hours consistent with on-site vendor recommendations. For students with an IP level of 3.0 or greater, the matrix provided indicates that usage at these recommended levels is capable of yielding at least 1.0 years of growth. For students with an IP level of 4.5 or greater, the matrix indicates that usage at these recommended levels is capable of yielding 1.5 years of growth. Data for this research spans 5 years beginning with the 2010-2011 school year.

**Participants**

Each year the program has been available, students with disabilities have had access to the program contingent on campus implementation plans. Consequently, some students have received multiple years of program usage. There is limited research available (McKissick, 2016) to suggest that multiple years of program use might affect program effectiveness. Each student-year of program use, then, will be considered unaffected by use in previous years.

The State of Texas has developed a number of end-of-year high-stakes examinations for its students. Prior to 2012, students took the Texas Assessment of Knowledge and Skills (TAKS). Five versions of that test were available to students: TAKS, a grade-level assessment identical to that taken by non-disabled students; TAKS-Accommodated, a grade-level assessment with additional allowable accommodations not believed to influence the rigor of the assessment; TAKS-Modified, testing grade-level concepts using simplified vocabulary, reduced answer choices, and a simplified format; TAKS-Alternate, for students with severe cognitive disabilities interfering with administration of paper-and-pencil examinations; and LAT, for students requiring linguistic accommodations. Beginning in 2012, students took the State of Texas Assessment of Academic Readiness (STAAR). Four versions of the STAAR were originally available, mirroring the versions available with TAKS, with the exception of a STAAR-Accommodated version. The STAAR-Modified test was replaced during the 2014-2015 school year with the STAAR-Accommodated version, an online assessment utilizing virtual tools such as a highlighter and sticky notes. State testing expectations are considered annually as part of the development of Individualized Education Plan for each student with disabilities.

During the five years of SMM use in the district, 2,441 student-years of data were collected. Of these, 156 were removed because prior-year (baseline) or current-year state testing data included the Alternate or linguistically accommodated version of the state assessment. Some students were introduced to SMM but did not complete initial placement. The reporting of state testing data for the previous year was taken as evidence that the student began the year in the district, and reporting of state testing data for the year of SMM was taken as evidence that the student ended the year in the district. Thus, an additional 668 were removed for lack of current- or prior-year state test data or SMM usage data indicative of either lack of treatment exposure or limited use due to partial-year enrollment. An additional 15 student-years of data were removed because no special education eligibility could be verified. Of the resultant 1,603 student-years of data, 398 included current- and prior-year state testing data at the different levels (grade-level or modified). These were removed for lack of adequate techniques to compare scores between various levels of the state assessments. The resultant dataset included 1,204 student-years of data from 920 unique students. There were 673 students who used the program for one year, 210 in two different years, and 36 students in three different years.

**Materials and Procedure**
SMM was made available for all middle school campuses in the district for use with students with disabilities. Campuses assumed responsibility for implementation of the program, including which students would access the program during various times of the day. Students at most of the campuses were provided opportunities to use the program before and after school as time and access allowed. Students were also able to access the program from home. Campus plans have undergone revision and refinement in subsequent years, and some campuses have integrated SMM use as part of the curriculum for resource mathematics classes (McKissick, 2016). Variations in campus implementation plans have not changed the specific intervention, namely SMM.

The district provided two measures of student achievement. First, SMM cumulative usage reports by student for each year were reviewed. These reports included an initial placement score, a grade level placement identified by SMM based on an initial evaluation of student abilities. A final grade placement score was also included so that a measure of math achievement gain during program use could be calculated. Because students from multiple grade levels were to have their performance analyzed simultaneously, it was determined that a measure of previous learning was needed. It was expected that students beginning a grade level should have an initial placement score equal to that grade level, indicative of achieving one academic grade level for each prior year of school. Thus, an average rate of growth was calculated by dividing the initial placement score by the grade. Additionally, state testing results from the previous year were made available. As mentioned above, changes in state testing have been frequent. Though scaled scores were made available, changes in scales between test versions and across years have made comparisons nearly impossible. Using district means and standard deviations, these scores were transformed to *z*-scores by test type and year. The design for this study is modeled in the diagram below, where *O1* and *O2* represent state testing results and SMM grade placement results respectively:

| | | | |
|---|---|---|---|
| *NR* | $\{O_{1A}, O_{2A}\}$ | $X_{FULL\ (>20\ Hours)}$ | $\{O_{1B}, O_{2B}\}$ |
| *NR* | $\{O_{1A}, O_{2A}\}$ | $X_{LIMITED\ (15\text{-}20\ Hours)}$ | $\{O_{1B}, O_{2B}\}$ |
| *NR* | $\{O_{1A}, O_{2A}\}$ | $X_{LIMITED\ (10\text{-}15\ Hours)}$ | $\{O_{1B}, O_{2B}\}$ |
| *NR* | $\{O_{1A}, O_{2A}\}$ | $X_{LIMITED\ (5\text{-}10\ Hours)}$ | $\{O_{1B}, O_{2B}\}$ |
| *NR* | $\{O_{1A}, O_{2A}\}$ | $X_{LIMITED\ (0\text{-}5\ Hours)}$ | $\{O_{1B}, O_{2B}\}$ |

Students were classified by their level of program use. Group A used SMM for 0-5 hours during a year, Group B used the program for 5-10 hours during a year, Group C used the program for 10-15 hours, Group D used the program for 15-20 hours, and Group E used the program for more than 20 hours. Two revisions were made to the dataset. First, all students with an average rate of prior growth greater than 1.0 were removed. Though these 18 students had identified disabilities, it was not apparent that the disabilities had impacted their mathematics achievement. Second, it was determined that the unbounded upper end of Group E allowed for the inclusion of "super-users" who had accumulated well over 25 hours of program use (maximum use reported was 81.4 hours in a year). Consequently, Group E was amended to include students with 20-25 hours of program use, resulting in the exclusion of 194 "super-users." This resultant range coincides with vendor recommendations to the district regarding target usage levels.

A primary concern in the absence of random assignment is the establishment of between-group homogeneity. An analysis of variance identified no significant variations between groups regarding their prior year state testing performance, $F(4, 885) = 1.56$, $p = .1817$. Similar analyses were conducted between groups for all disability areas. A significant difference was found only among students with an intellectual disability, though the result may be due to a small number of students in the sample with that disability. An analysis of variance was conducted to determine if there were any differences between usage groups regarding the average rate of growth. Again, no statistically significant between-group differences were found, $F(4, 885) = 1.14$, $p = .3375$. Analyses for between-group differences in average rate of growth were conducted by disability area. Between-group differences existed for students with autism, likely due to small sample sizes. Summary information for theses analyses are provided in Table 4. Analyses of both variables were extended to grade, gender, ethnicity, and school year. All tests identified homogeneity of groups except for prior state testing in 2013 and average rate of growth in 2014. Both may indicate refinement of campus implementation plans, though it should also be noted that the state test changed from TAKS to STAAR for the 2013 school year. Based on these analyses, the usage groups demonstrate sufficient homogeneity to proceed with further analysis. Additional group description, including demographic information, is provided in Tables 5 and 6.

Table 4
*Tests for Group Homogeneity*
Dependent Variable 1: Average Rate of Growth Prior to SuccessMaker Use

| | | df | SS | MS | *F* | *p* |
|---|---|---|---|---|---|---|
| All Disabilities | Group | 4 | .0863 | .0216 | 1.137 | .3375 |
| | Error | 885 | 16.793 | .0190 | | |
| Autism | Group | 4 | .218 | .0544 | 3.025 | .0280 |
| | Error | 42 | .756 | .0180 | | |
| Emotional Disturbance | Group | 4 | .0257 | .0064 | .267 | .8979 |
| | Error | 44 | 1.059 | .0241 | | |
| Learning Disability | Group | 4 | .0565 | .0141 | .825 | .5093 |
| | Error | 540 | 9.244 | .0171 | | |
| Intellectual Disability | Group | 4 | .0272 | .0091 | 1.084 | .3861 |
| | Error | 15 | .1255 | .0084 | | |
| Other Health Impairment | Group | 4 | .0299 | .0075 | .388 | .8173 |
| | Error | 123 | 2.374 | .0192 | | |

Dependent Variable 2: State Testing *z*-Score for Year Before SuccessMaker Use

| | | df | SS | MS | *F* | *p* |
|---|---|---|---|---|---|---|
| All Disabilities | Group | 4 | 3.942 | .9855 | 1.565 | .1817 |
| | Error | 885 | 557.45 | .6299 | | |
| Autism | Group | 4 | 2.049 | .5121 | .991 | .4230 |
| | Error | 42 | 21.704 | .5168 | | |
| Emotional Disturbance | Group | 4 | 1.689 | .4223 | .618 | .6518 |
| | Error | 44 | 30.048 | .6829 | | |
| Learning Disability | Group | 4 | 2.996 | .7490 | 1.145 | .3346 |
| | Error | 540 | 353.35 | .6544 | | |
| Intellectual Disability | Group | 3 | 8.033 | 2.678 | 4.051 | .0270 |
| | Error | 15 | 9.916 | .6610 | | |
| Other Health Impairment | Group | 4 | 1.535 | .3838 | .6763 | .6097 |
| | Error | 123 | 69.799 | .5675 | | |

Table 5
*Usage Group Demographics*

| | Group A (0-5 hours) | Group B (5-10 hours) | Group C (10-15 hours) | Group D (15-20 hours) | Group E (20-25 hours) | Group F* (>25 hours) |
|---|---|---|---|---|---|---|
| *N* | 227 | 292 | 190 | 102 | 79 | 194 |
| Male/Female | 137 / 90 | 188 / 104 | 116 / 74 | 67 / 35 | 45 / 32 | 121 / 73 |
| Afr.-Amer. | 91 | 144 | 75 | 39 | 26 | 80 |
| Hispanic | 57 | 63 | 51 | 31 | 27 | 58 |
| White | 68 | 72 | 54 | 26 | 25 | 48 |
| Other | 11 | 13 | 10 | 6 | 1 | 8 |
| Autism | 16 | 10 | 8 | 7 | 6 | 15 |
| Emotional Disturbance | 16 | 11 | 13 | 6 | 3 | 7 |
| Learning Disabilitiy | 128 | 180 | 125 | 66 | 46 | 121 |
| Intellectual Disability | 5 | 7 | 5 | 2 | 0 | 6 |
| Other Health Impairment | 34 | 45 | 22 | 11 | 16 | 20 |
| Other Disabilities** | 7 | 7 | 3 | 4 | 1 | 6 |
| Multiple Disability Codes† | 21 | 32 | 14 | 6 | 7 | 19 |

*Notes*: *Group F was not included in the MANOVA and follow-up ANOVAs. **This category includes students who have auditory, visual, or orthopedic impairments. †Students may have disabilities in multiple areas. They are grouped separately here as the impact of multiple disabilities is not known.

Table 6
*Usage Statistics (Means and Standard Deviations) per Usage Group*

| | Group A (0-5 hours) | Group B (5-10 hours) | Group C (10-15 hours) | Group D (15-20 hours) | Group E (20-25 hours) | Group F* (>25 hours) |
|---|---|---|---|---|---|---|
| IP Level | 4.44 (1.02) | 4.27 (.94) | 4.36 (1.07) | 4.36 (.91) | 4.37 (.89) | 3.84 (1.04) |
| Avg. Growth Rate | .64 (.15) | .62 (.13) | .63 (.15) | .63 (.13) | .61 (.12) | .56 (.14) |
| Gain | .06 (.06) | .20 (.11) | .38 (.16) | .52 (.20) | .59 (.20) | 1.07 (.51) |
| Prior Year State Testing *z*-Score | -.64 (.84) | -.70 (.81) | -.76 (.73) | -.70 (.79) | -.51 (.76) | -.34 (.91) |
| Current Year State Testing *z*-Score | -.52 (.82) | -.60 (.84) | -.57 (.73) | -.48 (.80) | -.25 (.84) | -.23 (.91) |
| Accuracy | .62 (.16) | .65 (.09) | .65 (.08) | .64 (.09) | .63 (.07) | .62 (.07) |
| Questions per Session | 8.85 (6.71) | 14.62 (7.27) | 18.96 (8.17) | 20.93 (9.87) | 21.15 (9.52) | 23.38 (9.27) |
| Questions per Hour of Use | 37.76 (25.70) | 61.58 (26.76) | 75.08 (25.83) | 76.93 (23.52) | 75.85 (25.34) | 83.13 (27.96) |
| Session Length (in minutes) | 14.4 (4.2) | 14.4 (3.0) | 15.0 (3.6) | 16.2 (4.8) | 16.8 (5.4) | 16.8 (4.2) |

*Note*: Group F was not included in the MANOVA or follow-up ANOVAs.

## *Results*

Two outcome measures were identified that were consistent with the variables used to determine between-group equivalence. Prior rate of learning was subtracted from SMM-reported achievement gain to determine a change in learning rate. State testing scores from the year of program use and the year prior to program use were transformed to *z*-scores, and a *z*-score difference was derived by subtracting the two. The use of both measures was indicated by the dual expectations of program used – improvement in state testing performance and growth in student achievement rates.

Multivariate analysis of variance (MANOVA) was conducted to determine if student usage significantly affected these achievement measures. Attention was given to the assumptions of MANOVA prior to analysis. Assumptions regarding sample size, independence of observations, and types of variables used in the analysis appeared to be met. Analysis of univariate distributions for the dependent variables resulted in the removal of 49 outliers. Analysis of multivariate distributions, resulting in Mahalanobis distances, resulted in the removal of 53 outliers. Multivariate normality was determined by examination of the normality of each dependent variable, inspection of *Q-Q* plots, and review of residuals from a generalized linear model. For each usage level for each dependent variable, the Shapiro-Wilk *W* was not significant. These are provided in Table 7. The generalized linear model yielded a measure of overdispersion of 0.4328, the ratio of deviance to degrees of freedom. Overdispersion rates greater than 1 are problematic (Carruthers, Lewis, McCue, & Westley, 2008), so the assumption regarding multivariate normality was resolved. A comparison of linear and quadratic fit lines between the two dependent variables resulted in fractional increases to $R^2$, suggesting that a linear relationship between variables existed. The Levene statistic identified no variance concerns for the change in state testing *z*-scores. Comparison of group variances for the change in growth rate involved a comparison of the highest and lowest group variances. This yielded an $F_{MAX} = 2.048$, and the greatest ratio of sample sizes was 3.696. According to Tabachnick and Fidell (2001), "$F_{MAX}$ is the ratio of the largest cell variance to the smallest. If sample sizes are relatively equal (with a ratio of 4 to 1 or less for largest to smallest cell size, an $F_{MAX}$ as great as 10 is acceptable" (p. 80). To assess multicollinearity, the correlation between dependent variables was found to be low yet significant based on the sample size, $r = .082$ (95% CI [.016, .147]). The sample appears to meet all assumptions for the MANOVA. The MANOVA yielded a Wilks' $\Lambda = .5161$, $F(8, 1768) = 86.63$, $p < .0001$.

Table 7
*Shapiro-Wilk Values for DV Univariate Normality*

| | Change in Growth Rate | Change in State Testing $z$-Score |
|---|---|---|
| Group A (0-5 hours) $N = 227$ | $W = .9888$ $p = .0734$ | $W = .9908$ $p = .1619$ |
| Group B (5-10 hours) $N = 292$ | $W = .9908$ $p = .0655$ | $W = .9952$ $p = .5002$ |
| Group C (10-15 hours) $N = 190$ | $W = .9896$ $p = .1807$ | $W = .9917$ $p = .3443$ |
| Group D (15-20 hours) $N = 102$ | $W = .9852$ $p = .3128$ | $W = .9862$ $p = .3707$ |
| Group E (20-25 hours) $N = 79$ | $W = .9832$ $p = .3866$ | $W = .9832$ $p = .3878$ |

Univariate analysis of variance was conducted with each dependent variable. The analysis for change in state testing $z$-score was not significant, $F(4, 885) = 1.497$, $p = .2012$. Between groups $t$-tests found no usage groups to be statistically different for this outcome measure. The analysis of variance (ANOVA) for change in growth rate was significant, $F(4, 885) = 206.57$, $p < .0001$. All usage groups were statistically different from each other. Results for these analyses can be found in Table 8. The greatest change in growth rate was found for Group E, $\bar{x} = -.0213$ (95% CI [-.064, .021]). ANOVAs were also conducted to determine if there were any differences in both dependent variables for gender, ethnic, and disability groups; no group differences were found.

Table 8
*Results of ANOVAs for Each Outcome Measure for Groups A-E*

Change in Growth Rate

| Source | df | SS | MS | *F* | *p* |
|---|---|---|---|---|---|
| Usage Group | 4 | 30.457 | 7.614 | 206.57 | <.0001 |
| Error | 885 | 32.621 | .039 | | |
| Total | 889 | 63.078 | | | |

| Group | *N* | Mean | Lower 95%CI | Upper 95%CI |
|---|---|---|---|---|
| A | 227 | -.5819 | -.6069 | -.5569 |
| B | 292 | -.4230 | -.4451 | -.4010 |
| C | 190 | -.2514 | -.2788 | -.2241 |
| D | 102 | -.1010 | -.1383 | -.0637 |
| E | 79 | -.0213 | -.0637 | .0211 |

Change in State Testing *z*-Score

| Source | df | SS | MS | *F* | *p* |
|---|---|---|---|---|---|
| Usage Group | 4 | 2.538 | .634 | 1.497 | .2012 |
| Error | 885 | 375.153 | .424 | | |
| Total | 889 | 377.691 | | | |

Because the analysis of state testing *z*-scores was found to be not significant, attention was focused on the analysis of growth rate. All users, except for those with an average rate of growth before SMM use greater than 1.0, were considered for inclusion. This sample of 1186 included the 194 "super-users" excluded from previous analyses. In preparation for an ANOVA to determine if any variations existed between the six usage groups (previous five plus Group F, those who used the program for more than 25 hours) regarding a change in growth rate, the variable was analyzed for univariate normality. This resulted in the removal of 25 univariate outliers, resulting in a sample of 1161 student-years of usage. Subsequent Shapiro-Wilk *W* tests failed to confirm normality for 4 of the 6 groups on the dependent variable. A logarithmic transformation of the dependent variable was testing for univariate normality, and all groups demonstrated normality on the variable. A significant difference was found between groups, $F(5,1155) = 431.51$, $p < .0001$. Subsequent *t*-tests found significant differences ($p < .0001$) between all group pairings except Groups D and E (15-20 hours of use and 202-25 hours of use, respectively). Values for the means and confidence intervals of each group, converted into units of years change in growth rate, are provided in Table 9. The inclusion of previously excluded multivariate outliers resulted in minimal changes to the means for Groups A-D. The mean for Group E increased from the first to second ANOVA, though the 95% confidence interval still contains zero. The mean and confidence interval for Group F suggest that students with disabilities who use SMM for more than 25 hours are likely to realize significant changes in their rate of mathematics achievement.

Table 9
*Results of ANOVA for Change in Growth Rate for All Usage Groups*

| Source | df | SS | MS | *F* | *p* |
|---|---|---|---|---|---|
| Usage Group | 5 | 39.578 | 7.916 | 431.51 | <.0001 |
| Error | 1155 | 21.187 | .018 | | |
| Total | 1160 | 60.765 | | | |

| Group | *N* | Mean | Lower 95%CI | Upper 95%CI |
|---|---|---|---|---|
| A | 241 | -.5910 | -.6149 | -.5667 |
| B | 312 | -.4350 | -.4584 | -.4113 |
| C | 202 | -.2499 | -.2823 | -.2169 |
| D | 121 | -.0428 | -.0895 | .0051 |
| E | 94 | .0115 | -.0429 | .0674 |
| F | 191 | .4387 | -.3922 | .4860 |

*Note*: Means and confidence intervals have been converted from logarithmic values used in ANOVA to years of growth.

To determine if different student populations received differential benefit from program use, ANOVAs were conducted to determine variations existed within each usage group. No differences were found for gender or ethnicity groups. Small samples of students with intellectual disabilities and "other" impairments (not those with an OHI eligibility) were removed prior to analysis. No differences were found within usage groups to indicate differential impact of similar usage for students with different disabilities. ANOVAs were conducted across usage groups for each disability group. These analyses mirrored the combined ANOVA conducted above that indicated significant differences between all levels of usage. Results can be found in Table 10.

Table 10
*Results of ANOVA for Change in Growth Rate for Disability Groups*

| | Autism | Emotional Disturbance | Learning Disabilities | Other Health Impairment |
|---|---|---|---|---|
| *F* | $F(5,59) = 19.053$ | $F(5, 54) = 10.783$ | $F(5, 713) = 236.73$ | $F(5, 153) = 32.054$ |
| *p* | < .0001 | < .0001 | < .0001 | < .0001 |
| *N* | 65 | 60 | 719 | 159 |
| Usage Group | Mean (SD) | Mean (SD) | Mean (SD) | Mean (SD) |
| A | -.63<br>(-.86, -.40) | -.58<br>(-.73, -.43) | -.57<br>(-.61, -.53) | -.58<br>(-.69, -.46) |
| B | -.49<br>(-.77, -.20) | -.44<br>(-.63, -.26) | -.42<br>(-.46, -.38) | -.41<br>(-.51, -.31) |
| C | -.13<br>(-.43, .18) | -.11<br>(-.28, .06) | -.23<br>(-.28, -.19) | -31<br>(-.45, -.17) |
| D | -.09<br>(-.42, 22) | -.16<br>(-.40, .08) | -.04<br>(-.10, .02) | -.01<br>(-.19, .17) |
| E | .09<br>(-.25, .44) | -.05<br>(-.42, .32) | .06<br>(-.01, .13) | .02<br>(-.14, .18) |
| F | .84<br>(.61, 1.08) | .37<br>(.13, .61) | .43<br>(.38, .47) | .51<br>(.36, .67) |

Variation in usage patterns between campuses was identified. Fidelity of implementation has been identified as a reason why interventions fail (Mills & Ragan, 2000). A Chi-Square analysis of implementation variations between campuses, reflecting comparable number of students at each usage level, was significant, $\chi^2(44)= 245.77$, $p < .0001$. Students in Groups E and F, those who received the recommended usage and those who exceeded usage recommendations, were included in the same group for this analysis. Table 11 presents the percent of students from each campus that received or exceeded the recommended usage levels for each campus. The percentage of students in the current sample receiving or exceeding usage recommendations was 24.62%.

Table 11
*Campus Fidelity of Use*

| Campus | Total *N* for Campus | Percentage of Users Receiving or Exceeding Usage Recommendations |
|---|---|---|
| A | 89 | 14.61% |
| B | 145 | 31.03% |
| C | 154 | 20.13% |
| D | 86 | 17.44% |
| E | 72 | 45.83% |
| F | 82 | 54.88% |
| G | 139 | 17.99% |
| H | 25 | 16.00% |
| I | 147 | 13.61% |
| J | 80 | 53.75% |
| K | 76 | 1.32% |
| L | 91 | 18.68% |
| Total | 1186 | 24.62% |

*Note*: Totals cover the five years of usage for this review, and includes only students whose data was used in the analyses conducted.

Variations in usage patterns between usage level groups were also identified. Table 6 presents information regarding performance variables for each usage group. Accuracy is defined as the percent of exercises completed correctly. To achieve normality for this variable, 20 outliers were removed and an exponential transformation was applied. Six users were removed who had 0% accuracy (each attempted fewer than 12 questions), and an additional 5 users with 100% accuracy were removed (each attempted fewer than 5 questions). The resultant ANOVA identified a significant variation in accuracy between usage groups, $F(5, 1150) = 6.372$, $p < .0001$. Post-hoc *t*-testing identified that users in Group F had a significantly lower accuracy rate than users in Groups A-D (all $p < .0002$). Session length was calculated as the total usage time divided by the number of sessions (included in the SMM usage report). Attempts to normalize the variable were unsuccessful, so a non-parametric test was used to determine group differences. A Kruskal-Wallis analysis of variance by ranks found significant differences between groups on this variable ($H[5] = 98.107$, $p < .0001$).

Two measures of efficiency of use were identified. The number of questions per sessions provides a measure of the student's effort during each session of program use. To achieve normality for this variable, 16 outliers were removed and a square-root transformation was applied. Three users were removed who had 0% efficiency. All usage groups demonstrated normality except Group C (Shapiro-Wilk $W = .9832$, $p = .0154$), so interpretation of the resultant ANOVA should consider this normality concern. The ANOVA identified a significant variation in questions per session between usage groups, $F(5, 1161) = 126.52$, $p < .0001$. Post-hoc *t*-tests identified differences between all groups (all $p < .02$) except Groups D, E, and F. A second measure of efficiency, the number of questions per hour of use, was identified that removed the impact of session length differences between usage groups. Again, a square-root transformation was applied to achieve normality for each group level. Four outliers were removed, and three students with 0% efficiency were excluded from the analysis. The ANOVA identified a significant variation in the number of questions per hour between usage groups, $F(5, 1173) = 102.84$, $p < .0001$. Post-hoc *t*-tests identified difference between all pairings of Groups A and B with Groups C-F.

Each of these four performance variables was reviewed for differences between demographic groups. ANOVAs were conducted using the three transformed variables (accuracy, questions per session, and questions per hour), and a nonparametric test was conducted using session length. No differences for gender or ethnicity were found. Differences were found among disability groups for questions per session ($F[3, 999] = 3.475$, $p = .0156$) and session length ($H[3] = 9.626$, $p = .022$). Students with autism were found to answer more questions per session despite spending less time per session than students in other disability groups.

To determine the predictive capacity of these usage pattern variables regarding gain in achievement rates, a regression analysis was conducted. Since the amount of usage time has already been identified as having a significant impact on change in growth rates, this analysis was restricted to those students who had received or exceeded the usage recommendations ($N = 292$). A logarithmic transformation of time was required to achieve normality for this variable. The regression analysis identified time, accuracy, and questions per hour of program use as significant predictors of change in growth rate. Parameter estimates may be found in Table 12. A model including these three predictor variables accounted for 84% of the variance in student change in growth rate among students receiving or exceeding program usage recommendations ($R^2 = .8411$). Using the mean accuracy and mean number of questions per hour for these students, it was found that 25 hours of program use would result in growth rates commensurate with previous years of schooling. Increasing the use to 42 hours, holding the other two parameters constant, is predicted to yield a growth rate change of .5. This level of program use – nearly double the recommendations – may lead to closing the math achievement gap by half of a school year. To close the math achievement gap by a full school year, nearly 70 hours of program use is predicted to be necessary.

Table 12
*Regression Analysis Results*

| | Intercept | Time (Log-Transformed) | Accuracy (Exponential-Transformed) | Questions per Hour (Root-Transformed) |
|---|---|---|---|---|
| All students receiving or exceeding usage recommendations | -7.11* (-7.74, -6.48 | 1.02* (.95, 1.10) | 1.24* (1.04, 1.43) | .164* (.07, .258) |
| Students with Learning Disabilties | -4.81* (-5.34, -4.29) | .90* (.82, .98) | 1.99* (1.62, 2.36) | .009* (.004, .014) |
| Students with Other Health Impairments | -6.35* (-7.69, -5.00) | 1.07* (.85, 1.29) | 3.60* (2.42, 4.79) | .008** (.001, .016) |

*Note*: *Significant at $p < .001$. **Significant at $p < .03$

Regression analysis was also conducted for disability groups for those students receiving or exceeding usage recommendations. Small samples sizes prohibit generalizations for students with autism, emotional disturbances, and intellectual disabilities. Regression equations for students with learning disabilities and other health impairments (often, ADHD) identified the same parameters as significant. As the parameter estimates do not overlap, their differential impact may be of predictive value. Estimates for these parameters are also found in Table 12.

There are multiple ways to determine the effect size for the treatment used. When students are re-grouped dichotomously as to whether or not they received the treatment with fidelity, the impact on the outcome variable (logarithmic transformation in change in growth rate) is significant, $F$ (1, 1159) = 907.42, $p < .0001$, with an accompanying $R^2 = .439$. Using Kabacoff's (2014) formula below for using $R^2$ to find effect size, $f^2 = .78$.

$$f^2 = \frac{R^2}{1-R^2} \quad (1)$$

Cohen's (1988) recommendations for interpreting this statistic consider .35 to be a large effect. Using Cohen's (1988) formulae for converting between effect sizes, this effect size is equivalent to $d = 1.77$, large by Cohen's standards. Problematically, this calculation involves the use of SMM data for students who used the program sparingly (consider those with 0-5 hours of use). Information from SMM regarding yearly growth rates may be limited to a portion of the reporting year due to the limited use, therefore creating validity concerns regarding this interpretation.

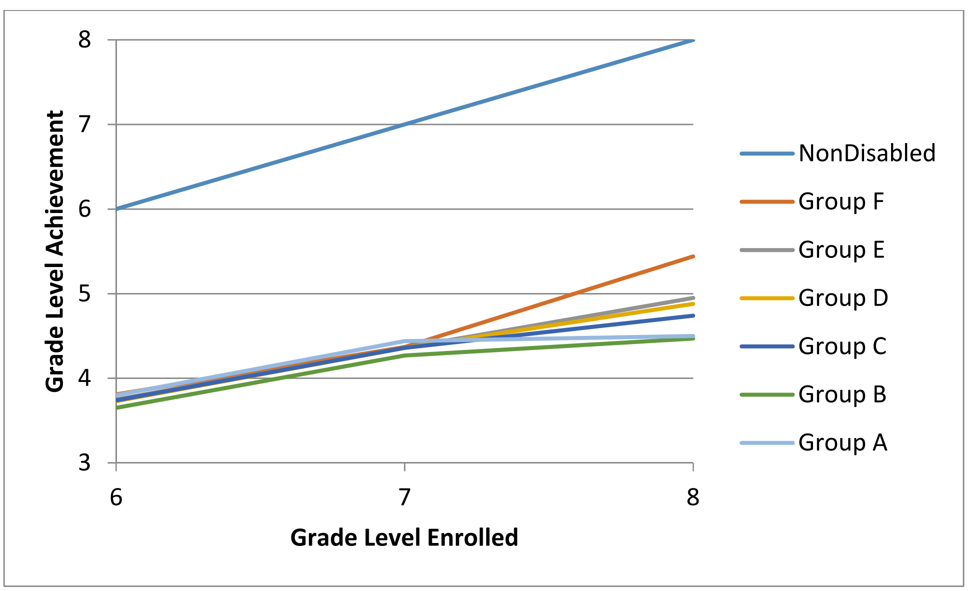
Alternately, students receiving the program with fidelity might have their rate of growth during treatment use compared to their rate of growth prior to SMM use. Students in Groups E and F, who met or exceeded usage recommendations ($N$ = 292), had a combined mean growth during treatment of .93 ($SD$ = .48). Their annual rate of growth prior to SMM use was .57 ($SD$ = .13). Using formulae (2) and (3) below from Ellis (2010), an effect size was found, Cohen's $d$ = 1.02. Cohen's (1988) benchmarks for evaluating effect sizes identify .80 as a large effect for this statistic. Similar comparisons for state testing performance utilize a prior mean $z$-score of -.387 ($SD$ = .884) and end-of-treatment $z$-score of -.240 ($SD$ = .889), yielding an insignificant effect size of $d$ = .01.

$$d = \frac{\bar{M}_1 - \bar{M}_2}{SD_{pooled}} \tag{2}$$

$$SD_{pooled} = \sqrt{\frac{(n_A - 1)SD_A^2 + (n_B - 1)SD_B^2}{n_A + n_B - 2}} \tag{3}$$

### *Discussion*

Regarding the effectiveness of SuccessMaker Mathematics for students with disabilities, the research conducted demonstrates the potential of the program for closing mathematics achievement gaps. The regression analyses identified that usage patterns regarding accuracy and efficiency (number of questions attempted per hour of program use), in addition to usage time, are useful predictors of changes in achievement growth rate. Though gender and ethnicity did not lead to group differences, variations between disability groups were present in various analyses. Figure 1 compares the changes in achievement growth rates for the six usage groups in this study to a hypothetical non-disabled student. Students are expected to experience one year of achievement growth for each year of school. Figure 1 illustrates that this has not historically happened for the disabled students using the program. Though the recommended use of SMM yields a learning trajectory similar to non-disabled students, much greater use would be needed to close the existing gaps.



*Notes*: The figure utilizes average growth rates and gains from Table 6. Data from 6th, 7th, and 8th grade students were consolidated into representative trend lines for 7th grade comparison. A hypothetical, non-disabled peer is provided as reference.

*Figure 1.* Learning trajectories of students with disabilities by usage group.

The use of outcome measures for this study present a variety of problems for interpreting the findings. State testing scores, the score of greatest concern to school districts, present significant comparison issues across years. Though equated scores may be useful for comparing across STAAR tests, no bridge was created to compare TAKS scores to STAAR scores. The issue is exponentially worse when addressing students with disabilities as the possible test versions and levels expands. This study has considered only those students whose state testing level (modified or on-level) remained constant from the previous year through the year of treatment. The use of *z*-scores for performance comparisons is less than desirable since students are compared to each other rather to an objective benchmark. Until the State of Texas provides a standardized and consistent measure of achievement, such poor comparison methods are likely to continue.

The consequence of poor state testing data is the need for measurement within SMM itself. Though the program provides an initial placement score, it is unable to assess student effort during the process. Consequently, students who are less motivated may intentionally perform poorly on the initial placement in an effort to meet a teacher's expectation for completion. It is believed that several students whose data was used in this study fall in this category of initial placement responding, though the large sample size and removal of outliers is believed to have reduced or eliminated their impact on analyses.

Further, use of treatment-provided achievement data as an outcome variable is not ideal. Identification and use of additional assessment instruments would be of assistance, and correlational analysis between those instruments and SMM would be useful. As with initial placement testing, performance on any other assessment instrument including state tests is subject to student motivational issues. A design that employs periodic evaluations of student motivation in addition to pre- and post-testing of achievement would improve upon these findings.

The quasi-experimental nature of this research also presents concerns. Though efforts were made to demonstrate homogeneity of usage groups on a host of factors, there is no good substitute for true random assignment. In the school setting, however, true randomization presents possible ethical and practical difficulties. Withholding access to a treatment believed to have benefit, especially for students with disabilities, may be ill-advised. Delaying access to treatment, as might be done in a design involving switching replications (Shadish, Cook, & Campbell, 2002), is difficult to implement for a year-long intervention. The use of a within-subjects design, as has been conducted here, may be necessary. Many interventions, such as SMM, are expensive purchases for school districts. In the absence of available funds or grants, researchers may be forced to utilize existing data. Forward-thinking districts are encouraged to develop an implementation plan that allows for appropriate data collection from the beginning to analyze program effectiveness.

This analysis considers effectiveness of SMM from a treatment dosage perspective. Students who received SMM with fidelity produced significantly higher mathematics achievement gains than students who did not receive the recommended usage of the treatment. When students who exceeded treatment usage recommendations are considered, those gains in achievement are even greater. Future research regarding SMM should consider implementing usage groups for greater usage levels than were considered for this project. Excessive use of the treatment was beyond the scope of this research. It is not yet known if use of SMM well beyond usage recommendations will result in continued linear growth or potential diminishing returns.

Though this paper has taken a pragmatist position, there is reason to believe that behaviorist instructional methods are helpful for students with disabilities. The behaviorist roots of SMM were reviewed above, and the effectiveness of the program for student with disabilities has been shown. This study did not investigate the use and perceptions of features more in line with cognitivist or constructivist theories. Instead, the repeated skill repetition and branching algorithms that serve as a foundation for skill presentation and assessment have yielded usage data consistent with this theoretical position. Further research that addresses the various components of the program is needed to determine what combined and individual effects these components have.

Previous research regarding SMM has included few studies in Texas. Most recently, Tucker (2008) found that SMM provided no benefit to 5$^{th}$ grade students using district passing rates as an outcome measure. This study has focused on the individual student, but has identified a similar lack of state testing differences following program use. Additionally, the current study has opted to address only those students with disabilities. Findings and conclusions from this study may not be generalizable to other student groups or school districts.

The need for effective remediation tools for students with disabilities is clear and ongoing. SuccessMaker has demonstrated an ability to assist struggling learners, but only if minimum usage recommendations are followed. Even then, these learners may not achieve learning gains commensurate with their non-disabled peers. Schools using SMM are encouraged to develop a clear plan for implementation that will allow students to meet targeted usage levels. Ongoing monitoring of student performance during program use is recommended so motivational issues discussed above may be addressed early. A discussion of implementation concerns is presented in McKissick (2016), though users are encouraged to identify the needs and target population for their campus.

## *References*


Bailey, G. D. (1992). Wanted: A road map for understanding integrated learning systems. *Educational Technology, 32*(9), 3-5.

Becker, H. J. (1992). Computer-based integrated learning systems in the elementary and middle grades: A critical review and synthesis of evaluation reports. *Journal of Educational Computing Research, 8*(1), 1-41.

Bransford, J. D., Sherwood, R. D., Hasselbring, T. S., Kinzer, C. K., & Williams, S. M. (1990). Anchored instruction: Why we need it and how technology can help. In D. Nix & R.

Spiro (Eds.), *Cognition, education, and multimedia: Exploring ideas in high technology* (pp. 115-142). Hillsdale, NJ: Erlbaum.

Burton, J., Moore, D., & Magliare, S. (2008). Behaviorism and instructional technology. In D. H. Jonassen & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology: A project of the Association for Educational Communications and Technology* (2nd Ed.; pp. 3-36). Mahwah, NJ: Erlbaum.

Cawley, J., Parmar, R., Yan, W., & Miller, J. (1998). Arithmetic computation performance of students with learning disabilities: Implications for curriculum. *Learning Disabilities Research and Practice, 13*, 68-74.

Cherryholmes, C. H. (1992). Notes on pragmatism and scientific realism. *Educational Researcher, 14*, 13-17.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cooley, M. L. (2007). *Teaching kids with mental health and learning disorders in the regular classroom: How to recognize, understand, and help challenged (and challenging) students succeed.* Minneapolis, MN: Free Spirit Publishing.

Crawford, A. N. (1970). A pilot study of computer assisted drill and practice in seventh grade remedial mathematics. *California Journal of Educational Research, 21*, 170-174.

Creswell, J. W. (2011). Controversies in mixed methods research. In. N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (4th ed., pp. 269-283). Thousand Oaks, CA: Sage.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: Sage.

Cruthirds, J., & Hanna, M. S. (1997). Programmed instruction and interactive multimedia: A third consideration. Retrieved from http://files.eric.ed.gov/fulltext/ED439464.pdf

Cummings, J. J., & Elkins, J. (1999). Lack of automaticity in the basic addition facts as a characteristic of arithmetic learning problems and instructional needs. *Mathematical Cognition, 5,* 149-180.

Delon, F. G. (1970). A field test of computer assisted instruction in first grade mathematics. *Educational Leadership, 28*, 170-180.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analysis, and the interpretation of research results*. Cambridge, UK: Cambridge University Press.

Gatti, G. G. (2009). *Pearson SuccessMaker math pilot study.* Pittsburgh, PA: Gatti Evaluation, Inc.

Gee, A. P. (2008). *An investigation of the impact of SuccessMaker on reading and math achievement at an elementary school* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (UMI: 3326592)

Graham, L., Bellert, A., Thomas, J., & Pegg, J., (2007). A basic skills intervention for middle school students with learning difficulties. *Journal of Learning Disabilities, 40*, 410-419.

Joyce, B., Weil, M., & Calhoun, E. (2009). *Models of teaching* (8th ed.). Upper Saddle River, NJ: Pearson.

Kabacoff, R. I. (2014). *Power analysis*. Retrieved from http://www.statmethods.net/stats/power.html

Kirk, V. C. (2003). *Investigation of the impact of integrated learning system use on mathematics achievement of elementary student* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (ProQuest Document ID: 305324110).

Kulik, C.-L. C., Kulik, J. A., & Cohen, P. A. (1980). Instructional technology and college teaching. *Teaching of Psychology, 7*(4), 199-205.
Kulik, C.-L. C., Schwalb, B. J., & Kulik, J. A. (1982). Programmed instruction in secondary education: A meta-analysis of evaluation findings. *The Journal of Educational Research, 75*(3), 133-138.
Kulik, J. A. (1994). Meta-analytic studies of findings on computer-based instruction. In E. L. Baker & H. F. O'Neil (Eds.), *Technology assessment in education and training,* (pp. 9-33). Hillsdale, NJ: Erlbaum.
Kulik, J. A. (2002). *School mathematics and science programs benefit from instructional technology* (Info Brief NSF 03-301). Washington, DC: National Science Foundation.
Manning, C. A. (2004). *The effect of the Math Concepts and Skills (MCS) computer program on standardized test scores at a middle school in east central Florida* (Doctoral dissertation). University of Central Florida, Orlando.
Martindale, T., Pearson, C., Curda, L., & Pilcher, J. (2005). Effects of an online instructional application on reading and mathematics standardized test scores. *Journal of Research on Technology in Education, 37*, 349-360.
McDonald, J., Yanchar, S., & Osguthorpe, R. (2005). Learning from programmed instruction: Examining implications for modern instructional technology. *Educational Technology Research and Development, 53*(2), 84-98.
McKissick, S. K. (2016). *Perceptions and obstacles encountered during SuccessMaker implementation* (Unpublished doctoral dissertation, Chapter 2). Texas A&M University, College Station, TX.
Mendelsohn, M. (1972). CAI in New York City: The slow learner in mathematics. *National Council of Teachers of Mathematics Yearbook*, 355-364.
Mills, S. C., & Ragan, T. J. (2000). A tool for analyzing implementation fidelity of an integrated learning system. *Educational Technology Research and Development, 48*(4), 21-41.
Mintz, K. S. (2000). *A comparison of computerized and traditional instruction in elementary mathematics* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses. (ProQuest Document ID: 304577773).
National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: NCTM.
Ormrod, J. (2014). *Human learning* (6th ed.). Essex, UK: Pearson.
Pearson Digital Learning (n.d.). *The Pearson timeline: Our history*. Retrieved from http://timeline.pearson.com
Pearson Digital Learning. (2002). *SuccessMaker evidence of effectiveness: Selected evaluation studies.* Retrieved from http://www.pearsoned.com/wp-content/uploads/dc4-successmaker-enterprise-evidence-of-effectiveness.pdf
Pearson Digital Learning. (2004). *Notes on SuccessMaker course levels, gain, and prescriptive scheduling*.
Pearson Education Inc. (2012, March). *SuccessMaker mathematics time/gain estimates for student scheduling: A guide for scheduling student sessions to achieve target scores.* Released publication from vendor provided by representative.
Pearson Education Inc. (2013). SuccessMaker 6 math reference guide. Retrieved from http://it.dadeschools.net/SuccessMakerNew/Teacher%20Resources/SuccessMaker_6_Math_Reference_Guide.pdf

Pellegrino, J. W., & Goldman, S. R. (1987). Information processing and elementary mathematics. *Journal of Learning Disabilities, 20,* 23-32.
Prince, J. D. (1969). *A practitioner's report results of two years of computer-assisted instruction in drill and practice mathematics*. McComb, MS: McComb Schools.
Ragosta, M. (1983). Computer-assisted instruction and compensatory education: A longitudinal analysis. *Machine-Mediated Learning, 1*(1), 97-127.
Richards, J. (1996). Negotiating the negotiation of meaning: Comments on Voigt (1992) and Saxe and Bermudez (1992). In L. P. Steffe, P. Nesher, P. Cobb, G. A. Goldin, & B. Greer (Eds.), *Theories of mathematical learning* (pp. 69-75). Mahwah, NJ: Erlbaum.
Schiro, M. S. (2013). *Curriculum theory: Conflicting visions and enduring concerns* (2nd ed.). Thousand Oaks, CA: Sage.
Schmidt, M., Weinstein, T., Niemiec, R., & Walberg, H. J. (1985). Computer-assisted instruction with exceptional children. *Journal of Special Education, 19*, 493-501.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.
Skinner, B. (1986). Programmed instruction revisited. *The Phi Delta Kappan, 68*(2), 103-110.
Slavin, R. E. (1987). A theory of school and classroom organization. *Educational Psychologist, 22*(2), 89-109.
Slavin, R. E., & Lake, C. (2008). Effective programs in elementary mathematics: A best-evidence synthesis. *Review of Educational Research, 78*(3), 427-515.
Suppes, P., & Morningstar, M. (1969). Computer-assisted instruction. *Science, 166*, 343-350.
Svoboda, D. S., Jones, A. L., van Vulpen, F., & Harrington, D. (2012). Programmed instruction. In J. Hattie & E. M. Andermann (Eds.), *International guide to student achievement* (392-395). New York, NY: Routledge.
Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.
Underwood, J., Cavendish, S., Dowling, S., Fogelman, K., & Lawson, T. (1996). Are integrated learning systems effective learning support tools? *Computers & Education, 26*(1), 33-40.
Van Dusen, L. M., & Worthen, B. R. (1995). Can integrated instructional technology transform the classroom? *Educational Leadership, 53*(2), 28-33.
Vockell, E. L., & Mihail, T. (1993). Behind computerized instruction for students with exceptionalities. *Teaching Exceptional Children, 25*(3), 38-43.
Vygotsky, L. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
Wendling, B. J., & Mather, N. (2009). *Essentials of evidence-based academic interventions*. Hoboken, NJ: John Wiley & Sons.
Wood, K. M. (2004). Effects of SuccessMaker Math on students with learning disabilities in inclusive and special education classrooms. *Journal of Teacher Initiated Research, 1*. Retrieved from http://www.otterbein.edu/Files/pdf/Education/JTIR/VolumeI/woodfinal.pdf
Zafiropoulou, M., & Karmba-Schina, C. (2005). Applying cognitive-behavioral interventions in Greek mainstream school settings: The case of learning difficulties. *Learning Disabilities: A Contemporary Journal, 3*(2), 29-48.

***Author Note***

Steve McKissick, Department of Educational Curriculum and Instruction, Texas A&M University.

The content of this document does not reflect any position or expression of the Killeen Independent School District, the KISD Board of Trustees, or the District's Administration. Correspondence concerning this article should be addressed to Steve McKissick, 1808 Bailey Dr., Copperas Cove, TX 76522. Email: mckissicks@tamu.edu