IMPLEMENTATION FIDELITY AND ATTAINMENT IN COMPUTERISED ASSESSMENT OF MATHEMATICS

Measuring the implementation fidelity (IF) or integrity of interventions is crucial, since without this a positive or negative outcome cannot be interpreted. Direct and indirect methods of IF measurement have been used in the past but tend to over-emphasize teacher behavior. This paper focuses on IF measured by student behavior collected through computers - an interesting alternative. Mathematics attainment was measured by the Star Maths test (a computerised item-banked adaptive norm-referenced test of mathematics). Implementation quality (IF) was measured by Accelerated Maths (AM) (an instruction-free personalized practice and progress-monitoring system that helps teachers monitor pupil progress in quantity, difficulty and mastery of mathematics skills). Attainment data was gathered in the UK on 20,103 students in 148 schools, and of these implementation data on n=6285. Attainment indices correlated well with each other but correlations with implementation indices were much more modest. Practice correlated best with attainment, followed by Objectives Mastery. There was a positive relationship between high implementation and high attainment, especially for Practice and Diagnostics, but students meeting the recommended implementation index levels were relatively few in number. Primary schools did better than secondary schools. Low socio-economic students did slightly better than the rest. Males and females were no different. High achievers were not better on implementation measures. Implementation quality is clearly a problem in the UK and steps should be taken to improve it, especially in secondary schools. However, overall students still scored above average on attainment.

The data were analyzed and this report prepared for Renaissance UK by Professor Keith Topping, Education, University of Dundee.

IMPLEMENTATION FIDELITY AND ATTAINMENT IN COMPUTERISED ASSESSMENT OF MATHEMATICS

Introduction

Since the emphasis has moved towards "evidence-based" interventions, measuring the quality of implementation has become an increasing preoccupation. Clearly, there is little point attempting to implement an evidence-based intervention and measure the outcomes if there is no parallel attempt to see whether the method has actually been implemented. Unless IF is assessed, in a circumstance of poor outcome we cannot know whether the program did not work or merely was not implemented properly, or both. Indeed, even in a circumstance of good outcome, we also cannot know whether the program actually worked and was responsible for the positive outcome.

Implementation fidelity (or integrity) (IF) was initially defined as the degree to which an intervention or treatment was implemented as planned, intended, or originally designed. However, this only specified the behavior of the interventionist, not that of the recipients of the intervention. By contrast, Schulte, Easton, and Parker (2009) included features related to the delivery of the intervention, how the intervention was received by the participants, and how the participants were able to use the learned skills in a natural environment. Schulte et al. (2009) among others espoused five elements: adherence to an intervention, exposure or dose, quality of delivery, participant responsiveness and program differentiation (the extent to which key factors in effectiveness are identified). Of course, the question then arises of which of these many indices are most related to outcome.

Measuring IF is not easy - researchers find that it is both complex and expensive. Indirect attempts which simply asked teachers whether they had implemented well were often found not to correlate with outcomes. Direct attempts which used observational methods (to avoid teacher subjectivity) were expensive (and consequently usable only on a small scale) and could still suffer from observer effects – what the teacher did when observed might not have been typical of what they did when not observed. Teacher behavior is the focus of much of the literature. Schulte et al.'s (2009) inclusion of participant responsiveness has been largely overlooked. There is also an issue about how often IF should be assessed, since many of the reports in the literature are of short-term interventions.

The Current Paper

Mathematics generates a large amount of data by computerised assessment. Data is also gathered over the course of a whole school year. This paper focuses on the effectiveness of differentiated practice in mathematics via the Accelerated Maths (AM) program and emphasizes student response rather than teacher behavior. The paper compares and contrasts five different kinds of implementation indicators of IF with growth in mathematics attainment on the Star Maths test. The study deploys measures of student response to counter-balance the existing over-emphasis on teacher behavior. In this paper both attainment and IF measures are completed locally but scored online centrally, and the results fed back locally, all by computer. This central scoring enables the collection of large samples of data. The present paper is a companion to a previous paper focused on reading (Topping, 2017).

2.0 Previous Research on IF in Mathematics

The present literature review interrogates previous research on IF in mathematics. An initial section explores research on implementation fidelity in mathematics for programs other than AM. Some of these describe research on indirect measures (self-reports completed subjectively by teachers and head teachers), while others focus on direct measures (completed by observation, although still far from "objective" given possible observer effects). After this, a further section briefly explores the outcome literature using systems such as Star Maths and AM. Then a final section explores in more detail such outcome tests used in conjunction with implementation measures.

2.1 Methodology of the Literature Review

The Social Sciences Citation Index (SSCI) and the Educational Research Information Centre (ERIC) were searched from 1995 to date (the terms implementation/treatment fidelity/integrity had little currency prior to this date). Search terms were “mathematics” AND “implementation fidelity” OR “implementation integrity” OR treatment fidelity” OR “treatment integrity”. The inclusion criteria specified relevance to the research questions and containing empirical data. Fifty-two hits resulted from the search of titles and abstracts. On reading the full text, a number of the papers still proved to be opinion pieces, reducing the items for the final literature review to 31.

2.2 Implementation Fidelity in Mathematics for Programs Other Than AM

A number of papers on IF in maths have disappointing findings - there was no evidence that the intervention was implemented as desired and no evidence of any improvement in attainment. An example is the evaluation of Classroom Assessment for Student Learning (CASL), a widely used professional development program (Randel, Apthorp, Beesley, Clark & Wang, 2016). Schools were randomly allocated CASL or regular professional development. Analysis of 67 schools and 9,596 students yielded no statistically significant impacts of CASL on student mathematics achievement as measured by the state-wide test. No statistically significant impacts were found on teachers' assessment practice.

More positively, Kinzie, Whittaker, McGuire, Lee, & Kilday (2015) evaluated the Research on Curriculum Design (RCD) model for pre-kindergarten mathematics and science curricula. Implementation spanned two years and involved iterative development and testing. A final test of the resulting curricula in eight pre-K classrooms yielded high-quality, high-fidelity teacher implementation, with teacher fidelity and curricular dosage predicting students' mathematics learning gains.

Unusually, Wolfe, Clements, Sarama, & Spitler (2013) focused on IF over time, examining the sustainability of teachers' implementation fidelity in a prekindergarten mathematics intervention, two years after external support ceased. Teachers continued to demonstrate high levels of fidelity to the underlying curriculum. Teachers with more experience demonstrated significantly faster rates of growth than their less experienced colleagues, and higher initial levels of fidelity to the whole group component.

Doabler, Nelson-Walker, Kosty, Baker, Smolkowski and Fien (2013) used a direct observation tool to systematically investigate the explicit instructional interactions that occurred during beginning mathematics instruction. However, many of the studied instructional variables were unrelated to student achievement.

Somewhat more positively, Crawford, Carpenter, Wilson, Schmeister and McDonald (2012) investigated fidelity and outcomes in a computer-based middle school mathematics curriculum for 485 students and 23 teachers from 11 public middle schools. Total time in intervention, concentration of time in intervention, direct observation of intervention fidelity and pre-test score were all found to be significant. However, fidelity to process was found to be nonsignificant.

Holstein (2012) examined teachers' implementation of a mathematical decision-making curriculum. Observations and teacher logs were coupled with interviews and surveys. Four out of six teachers were reasonably faithful to the program. Four types of implementers were identified: (a) "thorough piloting" teachers, (b) "adopting but adapting content" teachers, (c) "adopting but adapting pedagogy" teachers, and (d) "partial piloting" teachers. This typology may be useful elsewhere.

It is noteworthy that all but one of these studies focused on a short period of implementation. Few reported IF over a longer period such as a year. Even fewer reported IF indices available as a matter of course without additional effort as during the implementation of AM.

2.3 Outcome Literature Using Star Maths and Accelerated Maths

There is a good deal of outcome research on Star Maths and AM. The What Works Clearinghouse (2017) has summarized the primary level research (kindergarten through pre-algebra) in its own narrow way. Six studies met the WWC requirements, including 5,206 students in grades 2-9 in 223 classrooms across 27 states. The evidence for impact of AM on the mathematics test scores of students was medium to large.

Stanley (2011) studied AM with 624 third and fourth grade low socio-economic status high-ability students. Fourth graders who participated in AM did better on a maths attainment test than those who did not, but this was not true for third graders.

Ysseldyke and Betts (2010) examined maths curriculum data for nearly 2,000 classrooms assessing student achievement with Star Maths, some using AM and some not. Students who used AM with the following curricula outperformed peers who used only the curriculum: enVision Maths, Everyday Mathematics, Holt McDougal, Macmillan/McGraw-Hill, and Saxon Maths. AM resulted in significant positive effects irrespective of whether the maths curriculum was traditional or modern.

A randomized experiment with 28 at-risk high school juniors who used AM and a control group who participated in the school's typical maths classroom was conducted by Springer, Pugalee, and Algozzine (2007). Subsequently more experimental students (57%) were able to pass the state test than those in the control condition (14%).

A study in the UK by NFER (Rudd & Wade, 2006) involved matching 14 schools implementing AM with seven comparison schools. Comparisons over time (approximately an eight-month period) in primary schools showed maths attainment scores improved in two out of four intervention schools, and also in two comparison schools, although one comparison school saw a decline. In secondary both schools showed an improvement in scores, while scores in a comparison school did not improve to the same extent.

Also in the UK, Knock (2005) used AM in a daily 30-minute lunch-time maths club deploying AM with nine pupils and the same number of matched controls over two terms. The school had high socio-economic disadvantage (60% free school meals) and 30% ethnic minority background. The AM group improved on attainment tests three times more than the comparison group.

Grade 3–6 Title I students (n=870) from 47 schools in 24 US states were in a study by Ysseldyke, Betts, Thill, & Hannigan (2004). Students using AM significantly outperformed students who received only regular maths instruction. Another study with the same sample focused on gifted students (Ysseldyke, Tardrew, Betts, Thill, & Hannigan, 2004). Of the 100 gifted students in the sample, those who used Accelerated Maths advanced significantly more than those who did not.

Ysseldyke, Spicuzza, Kosciolek, and Boys (2003) examined the effects of AM on maths achievement. Sixty-eight percent of students were eligible for free lunch. Researchers assigned 160 students to classes with AM or without. Students in AM classrooms outperformed their controls.

Ysseldyke, Spicuzza, Kosciolek, Teelucksingh, et al. (2003) examined the effect of AM on the achievement of 881 students in grades 3–5 in four schools composed of 75% minority students and 67% free lunch eligibility. Students used AM with their regular curriculum, or just their regular curriculum on its own. Students at all ability levels who used AM demonstrated accelerated rates of improvement.

A quasi-experimental study by Gaeddert (2001) involved using AM with 103 students at a Kansas high school who were enrolled in pre-algebra, algebra, and geometry classes. The intervention group gained three times more than the control group.

Spicuzza, Ysseldyke, Lemkuil, Kosciolek, Boys, and Teelucksingh (2001) studied 198 fourth and fifth graders assigned to classrooms using AM or not. Most were eligible for free lunch. At all achievement levels, students in AM classrooms demonstrated more growth on Star Maths and the Northwest Achievement Levels Test than students in non-AM classrooms.

Teelucksingh, Ysseldyke, Spicuzza, and Ginsburg-Block (2001) studied 201 English learners (ELs) in grades 4-5 from four schools, comparing the maths performance of students using AM to a control group. AM students gained twice as much as the controls.

Anamourlis (2001) conducted a study of AM in Australia using 250 students from Years 3–7 from five schools throughout Australia over only five months. The teachers did not receive training and different classes spent an unequal time on AM. While AM and control groups showed similar gains in number, the AM group showed very large relative gains in areas of maths outside number.

Zumwalt (2001) studied 350 eighth-grade pre-algebra students in six schools. Ninety-four students received traditional instruction, 162 used AM, and 94 students used other computer-aided instruction software. The AM group out-performed the other two groups by a factor of 200%. In particular lower performing students benefited more from AM.

Clearly, then, there is no further need to show that AM has good outcomes. That has been established – although given that almost all the outcome research has been in the US, it is interesting to see if that is the case in another country. However, the extent to which different components of implementation contribute to these outcomes is another story, which we shall now explore.

2.4 Outcome and Implementation Fidelity in Accelerated Maths

AM offers a novel way of assessing IF, by taking indices of student responsiveness directly through computers (see Methodology section below for a fuller description of AM). However, relatively little of the previous literature has adopted this approach.

For instance, Lambert, Algozzine, and McGee (2014) categorized AM treatment classes in grades 2–5 at three Midwestern elementary schools into high- or low-implementation groups. Growth for the high-implementation group was significantly higher than for the low-implementation group, although both groups did better than non-AM controls.

Walker Driesel (2013) examined pre and post-test scores on Star Maths in relation to amount of classroom time dedicated to AM instruction. There was a strong correlation between the amount of time dedicated to AM and student performance, together with a decrease in the achievement gap when AM was implemented.

A sample of over 18,000 English Learners (ELs) and Native English Speakers (NESs) was examined by Lekwa (2012) to see if AM was implemented differently between the two groups, and test for differences in growth. Implementation of AM was a strong predictor of maths skill growth for both ELs and NESs, and ELs tended to receive slightly greater implementation of AM.

Burns, Klingbeil, and Ysseldyke (2010) examined data for 360 randomly selected schools in four states to compare three student groups: students using AM for over five years, for one to four years, and not at all. The groups using AM (>5 years and <5 years) had significantly more students scoring proficient on their states' high-stakes tests than the control group. Researchers also found that the achievement gap in control schools between primarily white and primarily ethnic-minority populations did not exist at the treatment schools.

To examine issues of variability, sustainability and IF, Bolt, Ysseldyke, and Patterson (2010) followed schools, teachers, and students for a second year following schoolwide implementation of AM. Students of teachers who implemented AM with greater fidelity experienced higher maths gains on standardized assessments than other students. Likewise, the relationship between AM use and achievement was significant at the student level.

The impact of Accelerated Maths with nearly 2,000 elementary students from eight schools and 100 classrooms in eight states was examined by Ysseldyke and Bolt (2007). Students were randomly assigned to AM with existing curriculum and control (only existing curriculum) groups. Students whose teachers used AM *as intended* demonstrated greater gains on two standardized tests (TerraNova and Star Maths) than students with limited or no implementation.

In another quasi-experimental study, Ysseldyke and Tardrew (2007) examined 2,200 students from 47 schools in 24 US states. Students using AM in grades 3–10 achieved maths gains from 7-18 percentile points higher than comparison students. In every grade and in Title I and free lunch programs, students in AM classes outperformed students not using the program. Students who closely followed AM best practice recommendations by scoring higher than 85% correct and completing more subskills made even greater gains. Low-, middle-, and high-achieving students showed consistent rates of gain for each maths objective mastered. The effects of the program were clearly a function of intervention integrity. This study was most like the present study.

In a multi-year analysis of maths achievement at 11 AM schools and 11 matched control schools, Nunnery and Ross (2007) found that both elementary and middle school students benefited from using AM, especially at high-implementing schools.

Holmes, Brown, and Algozzine (2006) examined the effectiveness of AM with 2,287 students from four elementary schools (two rural, two urban). One school in each area was either a high or low implementer of AM. Students in the two high-implementing schools outperformed students in the two low-implementing comparison schools overall (effect size [ES]=0.65), and separately in maths (ES=0.75), reading (ES=0.50), and language arts (ES=0.71).

A study in Germany examined AM with 22 fourth-, fifth-, and sixth-grade classrooms in 14 schools, matched with an approximately equal number of same-school, same-grade control classrooms that used their regular instructional methods. A total of more than 1,200 students participated (Lehman & Seeber, 2005). Over just four months achievement gains were unexpectedly high in both AM and control classrooms. Fifth-grade students using AM achieved an increase approximately twice that of the control group. In grades four and six, the AM and control students experienced similar levels of growth. Classrooms in which the AM program was used very extensively achieved the largest gains. Classrooms in which AM was not fully implemented achieved about the same gains as control classes.

3.0 Research Questions

Most previous studies of AM were in the USA and this was the first study in the UK for thirteen years, so it was of interest to see if US results could be replicated in another country.

1. Is better performance in implementation quality of Accelerated Maths associated with better mathematics outcomes on Star Maths?
2. Do primary/secondary status, gender, socio-economic status, ethnicity and mathematics ability influence these findings?

4.0 Method

4.1 Sample

The sample comprised all students in the UK for whom Star/AM results were available for the academic year (n = 20,103 in 75 elementary and 73 high schools). The years represented ranged from 2 through 13 and were approximately normally distributed, although year 7 was under-represented and year 8 was over-represented. The sample was generally representative of the whole of the UK, although Scotland was poorly represented (see the last section of the Results). However, the number of students for each analysis was generally large and is noted in the text.

4.2 Measures

4.2.1 Attainment – Star Maths. Star Maths is a computerised standardized (norm-referenced) adaptive item-banked mathematics test. It has mathematics questions with multiple-choice answers on a computer screen. It is standardized, i.e. any student's test responses are compared with the responses of many students of that age. The test is adaptive, i.e. it responds to the performance of each individual student. If the student succeeds on a question, harder questions are given. If the student fails, easier questions are given. This greatly reduces testing time and student stress. The test is also item-banked, i.e. it has multiple items at the same level. Consequently, students cannot copy from each other as no-one is doing the same test at the same time. This also enables the test to be taken frequently without practice effects. On completion feedback is available immediately to the teacher and/or pupil.

Star Maths has test-retest reliability of 0.93 in a US national sample of more than nine million tests. It also has internal consistency reliability of 0.97. More than 400 concurrent and predictive validity studies (correlations with many other measures of mathematics achievement) have been collected for Star Maths, involving 400,000 students. The average validity correlations range from 0.55 to 0.80. Correlations in that range are considered moderate to strong (Renaissance Learning, 2013).

4.2.2 Implementation fidelity - Accelerated Maths. Accelerated Maths (AM) is a personalized practice and progress-monitoring system that customizes maths practice assignments for students and helps teachers accurately and efficiently monitor student progress in quantity, difficulty and mastery of mathematics skills read. The program does not provide instruction – that is the role of the teacher. It does however provide practice in carefully differentiated skills for the student, a system of scoring and monitoring, and a system of feedback to the student and the teacher, thus minimising teacher administration time. Students are assigned to a series of practice activities on maths objectives, initially based on their entry score on Star Maths. AM automatically scores student work, and teachers can view reports that show performance. After reviewing student progress, teachers can adjust instruction for the entire class, for small groups of students struggling with similar objectives, or for individual students. Once students demonstrate mastery of a mathematical skill, AM automatically assigns new activities. AM currently includes content for KS1 to KS4. The equipment required is a class computer, printer and an Optical Mark Reader. Since students do their work on paper at their desks, and not at the computer, one set of equipment can serve the entire class.

Definitions of terms used in this paper are now offered.

4.3 Definitions

4.3.1 Achievement. Star Maths Scaled Score (SS) ranges from 0 to 1400 and spans years 1–13. It is based on the difficulty of the questions and the number of correct responses.

Star Student Growth Percentile (SGP) (Betebenner, 2011) is taken from SS scores on up to 3 tests within 18 months to give an indication of the student's growth trajectory. SGP is a norm-referenced percentile-based index ranging from 1-99. It indicates how exemplary a student's growth from one test window to another is relative to students in the same grade with a similar achievement history across the UK. SGP indicates past growth trajectory and predicts future growth trajectory. Because SGP is a mathematical manipulation, normal issues of reliability and validity do not apply, but issues of accuracy and precision do. Shang, VanIwaarden and Betebenner (2015) found that SGP tends to overestimate among students with higher prior achievement and underestimate among those with lower prior achievement, affecting 10% of students. Wright (2010) noted that SGPs correlated highly with value-added models but both under-estimated high-poverty classrooms, with SGP under-estimating least. The simulation-extrapolation method (SIMEX) was used to correct these anomalies.

4.3.2 Implementation. Average Percent Correct (APC) is the percent of correctness of the student's answers to the questions, in this case aggregated over a year.

Diagnostics is a test to gauge prior knowledge and identify gaps in skills. For APC on Diagnostics the recommended level of performance is >=85% correct or higher.

Practice is a test to assess whether the student is practicing the right skills at the appropriate level, time and pace. For APC on Practice the recommended level of performance is >=75% correct or higher.

Tests assess the students' level of mastery of the objective, once they have practiced skills sufficiently. For APC on Tests the recommended level of performance is >=85% correct or higher.

Objectives Mastered is a count of mastery of the concepts and subskills the students have been learning.

Review assesses the students' level of mastery of the objective, once they have practiced skills sufficiently. For APC on Review the recommended level of performance is >=80% correct or higher.

4.4 Data Analysis

Initially descriptive statistics were used to illuminate the data. Then Cohen's delta effect sizes (ES) were calculated to examine the importance of differences between variables. Effect Sizes of .01 were characterized as "very small" (Sawilowsky, 2009). Effect Sizes of .20 were characterized as "small", those of .50 as "moderate" and those of .80 as "large" (Cohen, 1988). Effect Sizes of 1.20 were characterized as "very large" and those of 2.0 as "huge" (Sawilowsky, 2009). In the interests of clarity and transparency, break points between these indices were added by the present author: 0.10 between .01 and .20, .35 between .20 and .50, .65 between .50 and .80, 1.00 between .80 and 1.20, and 1.60 between 1.20 and 2.0. Non-parametric correlation was carried out in some instances to examine relationships between variables.

5.0 Results

In these Results, we will first examine attainment on the Star Maths test, reporting gender, ethnicity, grade and socioeconomic status differences. Then we will examine AM implementation variables and attainment. Further analysis of primary-secondary (grade), gender, socio-economic, ethnicity and math ability differences in implementation is then reported.

On the attainment test Star Maths, a total of n=20,103 students had SS Gain scores (mean 75.47, sd 84.763), while SGP scores were available for n=19,841 (mean 52.10, sd 28.883). The SGP average score suggested that those students who took Star Maths performed overall at above the average level.

Gender in Attainment

Regarding Gender, on Star Maths n=1856 were marked as Unknown, n=820 were left blank, n=8765 were marked as Female and n=8662 were marked as Male. The mean for Males on SS Gain was 77.330 (sd=88.833), while that for Females was 74.760 (sd=80.381). Thus, Males appeared to do better than Females, but the ES was only .03 (very small) (the variance was large). On SGP, Males' average score was 52.81 (sd=30.410), while Females average score was 52.15 (sd=29.257). Again, Males did better than Females, but the ES was only .02 (very small). Thus, there was no substantial difference between Males and Females in attainment.

Variation by Grade in Attainment

Students with Star Maths scores were in Years 2 through 13 - a wide range. Table 1 shows the considerable variation in the numbers of students in each Year and also the considerable variation in their average scores by Year. In Year 2 students do not do very well, although numbers here are small. Thereafter the average scores on SS Gain seem to bounce up and down, but their average SGP scores continue to rise. This reached a peak of 62.28 in the first year of secondary school (Year 7). After this SS Gain and SGP sharply declines, although Year 8 has the highest number of students within it.

Free School Meal in Attainment

Students receiving Free School Meal or Not were recorded for 19,283 cases, 95.9% of the total number of students. Of these, 1,095 had Free School Meals, while 18,188 did not.

Free School Meals students scored an average of 81.79 (sd=86.936) on SS Gain and 52.71 (sd=29.635) on SGP. Not Free School Meals students scored an average of 75.26 (sd=84.346) on SS Gain and 52.14 (sd=29.896) on SGP. Thus, Free School Meal students did better than Not Free School Meal students on SS Gain but owing to the high variance the ES was only .08 (small). However, on SGP the ES was only .02 (very small).

Ethnicity in Attainment

In relation to Ethnicity, on Star Maths the main categories with substantial numbers were Asian, Black and White. Ethnicity was not recorded for 16,682 cases. Of those that were recorded, Whites were the most common (1999, 63%), Asians the next most common (897, 28%), and Blacks the least common (266, 8%).

Table 1
Variation by Year in SS Gain and SGP Gain

Year		SM_SSGain_PreToPost	SM_SGP_PreToPost
2	N	257	254
	Mean	113.58	45.91
	Std. Deviation	87.678	34.597
3	N	1519	1511
	Mean	129.99	49.98
	Std. Deviation	85.671	29.882
4	N	2382	2361
	Mean	113.72	51.32
	Std. Deviation	78.767	29.378
5	N	2514	2465
	Mean	95.13	51.06
	Std. Deviation	69.985	28.743
6	N	2801	2763
	Mean	96.03	60.03
	Std. Deviation	73.772	28.931
7	N	2165	2124
	Mean	79.24	62.28
	Std. Deviation	75.670	29.837
8	N	4415	4397
	Mean	45.35	48.38
	Std. Deviation	80.561	29.699
9	N	2660	2597
	Mean	35.83	48.24
	Std. Deviation	76.577	29.144
10	N	1060	1053
	Mean	33.91	47.68
	Std. Deviation	84.681	29.020
11	N	240	240
	Mean	21.35	41.56
	Std. Deviation	101.015	29.072
12	N	65	51
	Mean	33.83	40.04
	Std. Deviation	104.396	26.610
13	N	25	25
	Mean	44.56	47.96
	Std. Deviation	96.241	25.941
Total	N	20103	19841
	Mean	75.47	52.10
	Std. Deviation	84.763	29.883

Blacks had by far the highest scores on SS Gain (mean=104.630, sd=74.654) and SGP Gain (mean=57.870, sd=27.952). However, the number of blacks was small and may not be typical of their population. Asians had the next highest scores on SS Gain (mean=75.670, sd=75.568) and SGP (mean=53.760, sd=30.377). Whites had the lowest scores on SS Gain (mean=73.910, sd=87.658) and SGP (mean=53.460, sd=30.227). However, even the Whites' scores were above average on SGP. The differences in SS Gain between Blacks and Asians were substantial (ES .39, moderate), as were differences between Blacks and Whites (ES .38, moderate), but the difference between Asians and Whites was very small (ES .02). The differences in SGP between Blacks and Asians were less substantial (ES .14, small), as were differences between Blacks and Whites (ES .15, small), but the difference between Asians and Whites was very small (ES .01).

Implementation Data

The average SS Gain and SGP for the smaller sample was very similar to that for the larger sample. SS Gain mean = 71.17 (n=6285, sd= 81.508). SGP mean = 53.16 (n= 61.94, sd = 29.481). The ESs comparing large to small samples were .05 for SS Gain and .04 for SGP (both very small).

The number of students yielding implementation data on Objectives Mastered, APC_Practice, APC_Tests, APC_Review and APC_Diagnostics was much smaller than the number yielding attainment data. The number of students having APC_Practice scores was n=6,285, the largest number. Other implementation scores were based on n=5677 for APC_Tests, n=5530 for Objectives Mastered, n=4803 for APC_Review, and n=2370 for APC_Diagnostics.

Non-parametric correlations (Spearman's) were undertaken in view of the uncertain nature of the sampling of the students. SS Gain and SGP correlated highly at .823. However, other correlations between attainment and implementation were much more modest. The next highest set of correlations was between APC_Practice and attainment - .203 (SS Gain) and .241 (SGP). The next highest was Objectives Mastered at .133 (SS Gain) and .119 (SGP). These correlations are small and account for relatively little of the variance.

However, the implementation variables showed relatively high correlation with each other. APC_Practice correlated at .507 with Objectives Mastered, .381 with APC_Tests, .508 with APC_Review, and .541 with APC_Diagnostics. Objectives Mastered correlated at .362 with APC_Tests, .303 with APC_Review and .430 with APC_Diagnostics. All of these correlations were statistically significant on account of the relatively high numbers in the sample, but little weight should be given to this finding.

High Quality Implementation

Renaissance Learning recommends criteria indicating high quality implementation. These are APC_Practice >= 75%, APC_Tests>=85%, APC_Review>=80%, and APC_Diagnostic>=85%. Objectives Mastered does not lend itself to such recommendations since different objectives are not at the same level of difficulty, although obviously more is better at any level.

Only 403 students had APC_Practice scores >=75%. Their mean was .817 (sd=.055), with SS Gain 100.200 (SD=70.089) and SGP 64.600 (sd=28.319). APC_Practice<75% (n=5882) had a mean of .401 (sd=.158), with SS Gain 69.180 (sd=81.862) and SGP of 52.440 (sd=29.407). The APC_Practice <> mean difference was very substantial with ES 3.88 (huge). The SS Gain mean difference was quite substantial at ES .41 (moderate). The SGP mean difference was also quite substantial at ES .35 (moderate). Thus, there was considerable evidence validating the high-quality implementation criterion, although unfortunately only a small percentage of students (6%) came into this category.

Only 492 students had APC_Tests scores >=85%. Their mean was .912 (sd=.045), with SS Gain 86.950 (sd=81.857) and SGP Gain 59.460 (sd=29.524). APC_Tests <85%

(n=5184) had a mean of .466 (sd=.243), with SS Gain 69.730 (sd=80.848) and SGP Gain 53.320 (sd=29.285). The APC_Tests <> mean difference was very substantial with ES 3.10 (huge). The SS Gain mean difference was quite substantial at ES .25 (small). The SGP mean difference was quite substantial at ES .21 (small). Thus, there was some evidence validating the high-quality implementation criterion, although unfortunately only a small percentage of students (10%) came into this category.

Only 552 students had APC_Review scores>=80%. Their mean was .886 (sd=.066), with SS Gain 83.210 (75.678) and SGP 60.570 (sd=29.131). APC_Review <80% scores (n= 4251) had a mean of .493 (sd=.212), with SS Gain 68.960 (sd=82.592) and SGP 53.060 (sd=29.165). The APC_Review <> mean difference was very substantial with ES 2.83 (huge). The SS Gain mean difference was at ES .18 (small). The SGP mean difference was at ES .26 (small). Thus, there was some evidence validating the high-quality implementation criterion, although unfortunately only a small percentage of students (11%) came into this category.

Only 259 students had APC_Diagnostic scores>=85%. Their mean was .940 (sd=.050), with SS Gain 97.040 (sd=79.471) and SGP Gain 60.100 (sd=30.247). APC_Diagnostic <85% scores (n=2119) had a mean of .331 (sd=.260), with SS Gain 66.540 (sd=82.202) and SGP Gain 50.880 (sd=28.914). The APC_Diagnostics <> mean difference was very substantial with ES 3.93 (huge). The SS Gain mean difference was quite substantial at ES .38 (moderate). The SGP mean difference was also quite substantial at ES .31 (small). Thus, there was considerable evidence validating the high-quality implementation criterion, although unfortunately only a small percentage of students (11%) came into this category.

Gender in Relation to Implementation and Attainment

As Table 2 indicates, there were slightly more boys than girls in the implementation data. All differences in attainment and implementation were very small. Essentially there was little difference between boys and girls. Consequently, effect sizes were not calculated for this variable.

Year Level in Relation to Implementation and Attainment

For Primary n=2175 (29 Schools), mean SSGain = 103.510 (sd=73.925), mean SGP = 55.980 (sd=29.326), mean Objectives Mastered = 83.020 (sd=78.325), mean APC_Tests = .496 (sd=.301), mean APC_Practice = .477 (sd=.223), mean APC_Review = .526 (sd=.272), mean APC_Diagnostic = .516 (sd=.357).

For Secondary n=4110 (63 schools), mean SSGain = 54.05 (sd=80.137), mean SGP = 51.680 (sd=29.458), mean Objectives Mastered = 64.370 (sd=54.960), mean APC_Tests = .513 (sd=.244), mean APC_Practice = .402 (sd=.153), mean APC_Review = .543 (sd=.220), mean APC_Diagnostic = .343 (sd=.271).

For SS Gain, Primary was ahead of Secondary (ES .64 – moderate). For SGP, Primary was ahead of Secondary (ES .15 – small). For Objectives Mastered, Primary was ahead of Secondary (ES .64 – moderate). For APC_Practice, Primary was ahead of Secondary (ES .40 – moderate). For APC_Diagnostic, Primary was ahead of Secondary (ES .55 – moderate).

For APC_Tests, Secondary was ahead of Primary (ES .06, very small). For APC_Review, Secondary was ahead of Primary (ES .07 – very small).

However, inspection of attainment and implementation by Year revealed that the Primary/Secondary distinction was masking differences between Years (see Table 3). As with the whole attainment sample, year 7 (the first year of secondary education) showed good results, although more so in SGP than in SS Gain. Thereafter performance declined.

Table 2
Gender Differences in Attainment and Implementation

Female	SM_SSGain_ PreToPost	SM_SGP_ PreToPost	APC_Review	APC_Tests	APC_Practice	APC_Diagnostic	Objectives_ Mastered
N	2870	2848	2232	2604	2870	1137	2569
Mean	72.03	53.75	.5537	.4956	.4361	.3983	71.34
Std. Deviation	76.204	28.704	.22992	.25301	.18079	.30780	61.891
Male							
N	2981	2960	2310	2686	2981	1085	2634
Mean	70.75	53.25	.5351	.5217	.4306	.3978	72.37
Std. Deviation	85.250	30.027	.23374	.26051	.18214	.31430	66.643

Table 3

Year Level in Relation to Attainment and Implementation

Year		SM_SSGain_ _ PreToPost	SM_SGP_ PreToPost	Objectives_ Mastered	APC_Tests	APC_Practice	APC_Review	APC_Diagnostic
3	N	118	118	85	106	118	74	37
	Mean	114.94	38.68	42.44	.3331	.2624	.3920	.1351
	Std. Deviation	79.577	28.498	42.959	.30615	.16029	.24945	.30187
4	N	483	483	377	409	483	332	136
	Mean	116.54	52.44	87.94	.4702	.4491	.4645	.4697
	Std. Deviation	76.400	28.885	101.105	.31083	.22247	.26231	.38969
5	N	632	600	565	529	632	393	278
	Mean	94.99	51.31	86.05	.4631	.5269	.5003	.6182
	Std. Deviation	66.688	27.809	73.673	.31326	.24000	.27476	.32469
6	N	942	930	803	781	942	632	290
	Mean	101.12	63.03	82.87	.5309	.4837	.5909	.4883
	Std. Deviation	75.552	28.803	70.678	.27608	.20073	.26157	.33609
7	N	654	641	587	564	654	498	170
	Mean	80.33	63.12	67.45	.5250	.4873	.6026	.4924
	Std. Deviation	76.755	29.718	67.184	.25944	.19962	.27352	.37102
8	N	1504	1502	1367	1447	1504	1225	532
	Mean	59.04	50.92	64.97	.5394	.3973	.5444	.3534
	Std. Deviation	74.966	28.979	54.037	.23747	.13792	.20651	.24204
9	N	1241	1215	1070	1155	1241	1005	506
	Mean	43.59	48.45	57.30	.4857	.3680	.5345	.2926
	Std. Deviation	78.144	28.628	49.526	.24825	.13436	.20950	.26103

10	N	597	591	569	579	597	548	362
	Mean	37.60	49.36	68.11	.4793	.3783	.4930	.3177
	Std. Deviation	85.591	29.354	43.502	.22772	.12746	.19821	.23852
11	N	58	58	54	54	58	45	34
	Mean	29.55	40.83	105.81	.5594	.4534	.5471	.3506
	Std. Deviation	114.430	28.868	96.008	.23231	.17685	.25063	.28941
12	N	32	32	28	28	32	28	13
	Mean	45.13	45.28	77.32	.4889	.4278	.5854	.4600
	Std. Deviation	108.335	27.917	68.062	.24523	.16046	.20518	.32254
13	N	24	24	24	24	24	23	20
	Mean	46.33	49.21	72.71	.6579	.5446	.6665	.4335
	Std. Deviation	97.893	25.721	50.726	.22462	.11321	.19754	.20095
Total	N	6285	6194	5529	5676	6285	4803	2378
	Mean	71.17	53.16	70.54	.5043	.4278	.5380	.3969
	Std. Deviation	81.508	29.481	64.245	.26391	.18421	.23676	.31072

Free School Meals in Relation to Implementation and Attainment

Table 4 shows that only 554 students (9%) yielding attainment and implementation data had Free School Meals. Free School Meals students were a little below Not Free School Mealss students on both attainment measures, but the difference was very small. ES for SS Gain was .04 (very small), for SGP .01 (very small).

On Objectives Mastered, Free School Meal students were at almost the same level as Not Free School Meal students (ES<.01 – very small). Not Free School Meal students were ahead on three of the four APC measures (ES APC_Tests .12 small, ES APC_Practice .15 small, ES APC_Diagnostic .32 small). However. Free School Meal students were ahead on APC_Review – ES .04 very small).

Ethnicity in Relation to Implementation and Attainment

Table 5 shows the differences between ethnic groups on the variables of attainment and implementation. Students with a known ethnic origin numbered 1064, but this was only 16% of the total. Great caution is therefore needed in interpreting these figures.

In attainment, Black children were best by a considerable margin. Next were Asians, followed by Whites lagging a long way behind. In SS Gain, effect sizes Black/Asian were 0.25 (small), Black/White .67 (large) and Asian/White .41 (moderate). In SGP, effect sizes Black/Asian were 0.20 (small), Black/White .47 (moderate) and Asian/White .28 (small).

However, on Objectives Mastered, Asians did best, followed at a considerable distance by Whites, followed again at a considerable distance by Blacks. ESs were substantial: Asian/White .62 (moderate), Black/Asian .95 (large), Black/White .47 (moderate). Consequently, it is possible that Blacks achieved high attainment while only marginally engaging with Accelerated Maths – or perhaps they engaged with fewer objectives at an overall higher level of difficulty.

Table 4

Free School Meal in Relation to Attainment and Implementation

		SM_SSGain_ PreToPost	SM_SGP_ PreToPost	Objectives_ Mastered	APC_Tests	APC_Practice	APC_Review	APC_Diagnostic
No Free School Meal	N	5644	5614	5004	5099	5644	4339	2103
	Mean	71.52	53.19	70.88	.5058	.4313	.5384	.4033
	Std. Deviation	81.207	29.488	63.349	.26234	.18351	.23466	.31074
Free School Meal	N	554	541	440	491	554	384	242
	Mean	68.06	52.80	70.71	.4745	.4031	.5468	.3479
	Std. Deviation	85.423	29.820	74.272	.27165	.19273	.25553	.30517
Total	N	6198	6155	5444	5590	6198	4723	2345
	Mean	71.21	53.16	70.86	.5031	.4288	.5391	.3976
	Std. Deviation	81.591	29.515	64.293	.26329	.18451	.23641	.31057

Table 5
Ethnicity in Relation to Attainment and Implementation

Ethnicity		SM_SSGain_ PreToPost	SM_SGP_ PreToPost	Objectives_ Mastered	APC_Tests	APC_Practice	APC_Review	APC_Diagnostic
Unknown	N	5221	5169	4574	4662	5221	3958	1861
	Mean	71.93	53.15	67.93	.5030	.4244	.5419	.4014
	Std. Deviation	81.115	29.475	60.768	.26542	.18177	.23339	.31107
Asian	N	224	200	213	222	224	200	146
	Mean	84.40	57.25	127.68	.5323	.5604	.5734	.4692
	Std. Deviation	76.957	27.954	103.122	.26055	.21226	.26337	.31635
Black	N	109	107	77	95	109	61	17
	Mean	103.26	62.73	46.29	.3838	.4320	.4749	.3641
	Std. Deviation	71.647	28.063	68.240	.31429	.19805	.35303	.46453
Multiple	N	62	61	45	53	62	37	17
	Mean	91.42	57.44	49.93	.3804	.4287	.4757	.3565
	Std. Deviation	67.386	26.016	64.626	.27606	.20669	.32845	.35055
Other	N	51	50	39	44	51	36	13
	Mean	77.45	60.48	47.95	.4248	.3804	.3881	.2092
	Std. Deviation	106.799	30.239	53.759	.27288	.21462	.29822	.29056
White	N	618	607	581	600	618	511	324
	Mean	51.67	49.22	76.51	.5395	.4108	.5166	.3496
	Std. Deviation	83.122	29.905	61.534	.23134	.16741	.21520	.28667
Total	N	6285	6194	5529	5676	6285	4803	2378
	Mean	71.17	53.16	70.54	.5043	.4278	.5380	.3969
	Std. Deviation	81.508	29.481	64.245	.26391	.18421	.23676	.31072

On APC_Tests, Whites did best, followed by Asians and then Blacks. The difference between Whites and Blacks was ES .55 (moderate). Asians were also ahead of Blacks by ES .51 (moderate). The difference between Asians and Whites was ES .03 (very small).

However, on APC_Practice, Asians came first, followed by Blacks, who were followed by Whites. The difference between Asians and Blacks was ES .62. The difference between Asians and Whites was ES .78. However, the difference between Blacks and Whites was small, ES .11.

On APC_Review, again Asians came first, followed by Whites, followed by Blacks. The difference between Asians and Blacks was ES .32 (small). The difference between Whites and Blacks was ES .17 (small). The difference between Asians and Whites was ES .23 (small).

On APC_Diagnostics, again Asians came first, followed by Blacks and then by Whites. The difference between Asians and Blacks was ES .27 (small). The difference between Asians and Whites was ES .39 (moderate). The difference between Blacks and Whites was ES .04 (very small).

Whites did best on APC_Practice and came second on APC_Review but were last on APC_Practice and APC_Diagnostics. Especially given the low numbers in the sample, it is somewhat speculative to interpret these findings.

High/Low Ability in Relation to Attainment and Implementation

In relation to ability in mathematics, we could assume that SGP would indicate this with a mean of 100 and a standard deviation of 15. Thus, high achievers could be seen as students with SGP at or above 115, while low achievers could be seen as those with SGP at or below 85. We proceeded to compare high with low achievers but found that there were no high achievers. Consequently, we compared low achievers with the average for all students (Table 6).

Table 6
Differences by Low vs. Average Ability

Low Ability	SM_SSGain_ PreToPost	SM_SGP_ PreToPost	Objectives_ Mastered	APC_Tests	APC_Practice	APC_Review	APC_Diagnostic
N	5018	5018	4348	4502	5018	3808	1916
Mean	49.97	43.73	66.66	.4884	.4099	.5286	.3784
Std. Deviation	70.933	24.504	59.731	.26316	.17889	.23141	.30193
Average Ability							
N	6285	6194	5529	5676	6285	4803	2378
Mean	71.17	53.16	70.54	.5043	.4278	.5380	.3969
Std. Deviation	81.508	29.481	64.245	.26391	.18421	.23676	.31072

Unsurprisingly, average achievers did better on SS Gain (ES .28 - small) and SGP (ES .35 - moderate). However, there was little difference between average and low achievers on implementation: Objectives Mastered ES .06, APC_Tests ES .06, APC_Review ES .04, APC_Diagnostic ES .06 – all very small.

Region in Relation to Attainment and Implementation

We undertook an analysis by region to ascertain whether Accelerated Math schools were representative of the whole of the United Kingdom. In general, this was true, except Scotland was very poorly represented. The North of England had 1510 students, the Midlands 2069 students (boosted by a large number in Worcestershire) and the South 2100 students. Northern Ireland had 520 students but Wales only 46.

We then looked at the regions which had large numbers of students (>270) and arranged them in order of size of SS Gain (see Table 7). Staffordshire is only fourth on this list but has the highest SGP. West Yorkshire is second on the list but has a low SGP. The Regions highest on the list tend to have lower scores on APC_Tests. West Yorkshire is very low on APC_Practice. It is difficult to see a pattern in the other implementation indicators (APC_Review and APC_Diagnostic).

Table 7

Attainment and Implementation for Regions with Students n>=270

Region	SS_Gain	SGP	Objectives Mastered	APC_Tests	APC_Practice	APC_Review	APC_Diagnostic
Kent	106.29	60.10	103.84	.457	.624	.521	.533
West Yorkshire	100.60	49.15	60.83	.458	.365	.469	.236
London	97.76	61.26	89.40	.487	.502	.483	.539
Staffordshire	83.44	64.41	72.55	.500	.457	.584	.325
County Antrim	79.22	53.74	86.03	.520	.565	.610	.777
Essex	53.81	53.45	62.88	.583	.436	.508	.224
Worcestershire	53.00	50.23	53.09	.511	.375	.666	.165
Lancashire	43.58	46.73	63.53	.506	.329	.508	.284

6.0 Discussion

6.1 Summary

In attainment for the larger sample (n=20,103), the average SGP was 52.10, suggesting overall student performance was above average. Males and females had similar attainment. Variation over years showed a peak in the first year of secondary school, with sharp declines thereafter. Low socio-economic students actually did better than the rest of the sample, although effect sizes were small. Ethnicity was not known for many students, but where it was known Black students outperformed Asian and White students, although even White students still performed at above average levels.

Regarding attainment *together with* implementation indices, a much smaller sample was available (n=6285). Primary schools were about half as numerous as secondary schools and had about half the numbers of students participating – there was no evidence of larger numbers of participating students in secondary schools. SS Gain and SGP levels did not differ from the large sample and correlated highly. However, correlations with implementation indices were much more modest (maximum .241, accounting for relatively little of the variance). APC_Practice was the implementation index that correlated best with attainment, followed by Objectives Mastered. However, implementation variables correlated quite well with each other (maximum .508).

Considering high quality implementation, recommended levels are APC_Practice >=75%, APC_Tests>=85%, APC_Review>=80% and APC_Diagnostic>=85%. However, only 403 (6%) of students achieved this level for APC_Practice. Their ES for implementation was 3.88 (huge) and for SS Gain .41 (moderate) and SGP .35 (moderate). Only 492 (10%) of students achieved the recommended level for APC_Tests. Their ES for implementation was 3.10 (huge) and for SS Gain .25 (small) and SGP .21 (small). Only 552 (11%) of students achieved the recommended level for APC_Review. Their ES for implementation was 2.83

(huge) and for SS Gain .18 (small) and SGP .26 (small). Only 259 (11%) of students achieved the recommended level for APC_Diagnostics. Their ES for implementation was 3.93 (huge) and for SS Gain .38 (moderate) and SGP .31 (small). Thus, there was good evidence of a positive relationship between high implementation and high attainment, especially for APC_Practice and APC_Diagnostics.

Considering attainment and implementation data, no difference was found between males and females.

Regarding school sector, Primary schools did better than Secondary schools in attainment (SS Gain ES .64 moderate, SGP .15 small). There were also major differences on some implementation measures, namely Objectives Mastery (ES .64 moderate), APC_Practice (ES .40 moderate) and APC_Diagnostics (ES .55 moderate). Primary and Secondary were similar on the other two implementation measures. By Year, again Year 7 had good results, with a sharp decline thereafter.

Attainment and implementation data were related to students having Free School Meal or not. There was little difference between the groups. Not Free School Meal students were slightly ahead on three implementation indices and Free School Meal students were slightly ahead on attainment and APC_Review.

Concerning ethnicity, the numbers were small and great caution is needed in interpretation. However, in attainment Blacks were generally ahead of Asians who were ahead of Whites. However, Blacks did not perform well with regard to implementation. Asians were best on Objectives Mastered and Blacks the worst, perhaps suggesting Blacks were not fully engaged with Accelerated Math. On APC_Tests Whites were top, but they were worst on APC_Practice and APC_Diagnostics.

Regarding the difference between students of High and Low ability, unsurprisingly the High achievers were better on attainment. However, there was little difference on implementation measures.

Finally, examination of Regional variations showed that the sample was representative of the United Kingdom except for Scotland. Regions high on the list for attainment tended to have lower APC_Test scores. Beyond this there was no real pattern.

6.2 Connection to Previous Literature

As noted in the review of previous literature, many studies found implementation quality related to attainment. However, assessment of quality of implementation generally had not been in terms of the implementation indices used here, but in a much broader categorization. An exception to this is the study of Ysseldyke and Tardrew (2007), who noted that students who closely followed AM implementation recommendations by scoring higher than 85% correct and completing more subskills (Objectives Mastered) made the largest gains. This study is most like the present study, except it took place in a different country and several years ago, and did not use all five of the implementation indices.

6.3 Limitations and Strengths

The present study had a number of limitations, as well as a number of advantages. The principal advantage was the large sample size. This led to a de-emphasis on statistical significance. The sample was representative of years from 2-13 and of the United Kingdom, except for Scotland. SGP tends to under-estimate schools in socio-economically disadvantaged areas and over-estimate schools in advantaged areas, so when interpreting the tables some flexibility will be required. Outcome variables and implementation variables were not highly correlated with each other because the majority of students were performing below the recommended thresholds on the implementation criteria.

6.4 Implications for Practice, Policy and Future Research

6.4.1 Practice. Teachers should strive to maximize implementation levels of what appear to be the major determinants of higher outcomes– APC_Practice, APC_Diagnostics and Objectives Mastered. Of course, teachers are working indirectly with individual students who generate the data, so much of their work will involve explaining to students and subsequently coaching them. At a systemic level, when teachers evaluate the success of AM in their schools, they should carefully consider the evidence on these three key indicators of IF as well as the level of student outcomes and strive to increase them.

6.5.2 Policy. Policy-makers (including school inspectors) at local and national level should carefully consider the evidence on these three key indicators of IF as well as the level of attainment. Policy-makers need to be sharply aware that trials without accompanying reliable evidence of implementation integrity are of little value and should not be over-interpreted. The advice that they give to teachers should reflect this caution. They should consider providing relevant professional development opportunities to teachers and schools.

6.5.3 Future research. *S*hould studies similar to this be undertaken, in whatever country, it would be useful to investigate the three key implementation variables and their relationship to attainment. A further study of indirect, direct and computerised methods of establishing IF in mathematics with the same pupils would be highly valuable. As in this study the level of adequate implementation was so low, conducting further analysis of the relationship between implementation and attainment with these data would be of limited usefulness. If the level of implementation was higher, regression analyses linking implementation with attainment and allowing the prediction of attainment from level of implementation would be possible.

7.0 Conclusion

Thus, in relation to the research questions, we did find (RQ1) that better performance in the implementation quality of Accelerated Math was correlated with better mathematics outcomes on Star Maths, particularly for Average Percent Correct Practice and for Objectives Mastered. We also found that high implementation was associated with high attainment, especially for APC Practice and APC Diagnostics. Nonetheless, on average AM students performed at above average levels.

With regard to the second research question (RQ2), in general primary schools did better than secondary schools on both implementation and attainment. There was little difference between the genders on implementation or attainment. Free School Meal students did a little better on attainment and on APC_Review, while Not Free School Meal students did a little better on the other implementation indices. Regarding Ethnicity, data available was a small subset and must be regarded cautiously, but on attainment Blacks did better than Asians who did better than Whites. However, on implementation, Asians did best and Blacks did worst. High ability students were unsurprisingly better on attainment, but about the same as Low Ability students on implementation.

Indirect, direct and computerised student-response measures of IF all had some problems in predicting pupil outcome. It is suggested that future research needs to triangulate indirect, direct and computerised student-response measures with the same students over a period of at least a year, to establish which combination might be the most predictive in the longer run.

Computerised student response measures are not yet available in many other areas of the curriculum. Computerised methods of assessing teacher behavior seem to be some way in the future. A much larger portion of research resource needs to be devoted to establishing satisfactory multi-component IF measures. Researchers and research organizations interested in evidence-based interventions need to give much closer attention to the issue of IF.

References

Anamourlis, A. (2001). The impact of the Accelerated Maths pilot program in Australia. Boxhill: Renaissance Learning Australia.

Betebenner, D. W. (2011). *A technical overview of the student growth percentile methodology: Student growth percentiles and percentile growth projections/trajectories*. Dover, New Hampshire: The National Center for the Improvement of Educational Assessment. Retrieved from http://www.nj.gov/education/njsmart/performance/SGP_Technical_Overview.pdf [14 June 2016].

Bolt, D., Ysseldyke, J., & Patterson, M. (2010). Students, teachers, and schools as sources of variability, integrity, and sustainability in implementing progress monitoring. *School Psychology Review, 39*(4), 612–630.

Burns, M. K., Klingbeil, D. A., & Ysseldyke, J. (2010). The effects of technology enhanced formative evaluation on student performance on state accountability math tests. *Psychology in the Schools, 47*(6), 582–591.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York and London: Routledge. ISBN 1-134-74270-3.

Crawford, L., Carpenter, D. M., Wilson, M. T., Schmeister, M., & McDonald, M. (2012). Testing the relation between fidelity of implementation and student outcomes in math. *Assessment for Effective Intervention, 37*(4), 224-235.

Doabler, C. T., Nelson-Walker, N., Kosty, D., Baker, S. K., Smolkowski, K., & Fien, H. (2013). *Explicit instructional interactions: Observed stability and predictive validity during early literacy and beginning mathematics instruction.* Washington, DC: Society for Research on Educational Effectiveness. ERIC Number: ED563121. Available https://files.eric.ed.gov/fulltext/ED563121.pdf [12 June 2018].

Gaeddert, T. (2001). *Using Accelerated Math to enhance student achievement in high school mathematics courses.* Unpublished master's thesis. Friends University, Wichita, Kansas. ERIC Number: ED463177.

Holstein, K. A. (2012). *A characterization of teachers' implementations of a mathematical decision-making curriculum*. ProQuest LLC, Ph.D. Dissertation, North Carolina State University. ERIC Number: ED550966.

Kinzie, M. B., Whittaker, J. V., McGuire, P., Lee, Y. J., & Kilday, C. (2015). Research on curricular development for pre-kindergarten mathematics and science. *Teachers College Record, 117*(7), 136-157.

Holmes, C. T., Brown, C. L., & Algozzine, B. (2006). Promoting academic success for all students. *Academic Exchange Quarterly*, *10*(3), 141–147.

Lambert, R., Algozzine, B., & McGee, J. (2014). Effects of progress monitoring on math performance of at-risk students. *British Journal of Education, Society and Behavioural Science*, *4*(4), 527–540. Retrieved from http://www.journalrepository.org/media/journals/BJESBS_21/2014/Jan/Lambert442013BJESBS7259_1.pdf

Knock, D. J. (2005). Nottingham pupils improve mathematics achievement with Accelerated Maths. Wisconsin Rapids, WI: Renaissance Learning.

Lehman, R. H., & Seeber, S. (2005). Accelerated Mathematics in grades 4-6. Wisconsin Rapids, WI: Renaissance Learning.

Lekwa, A. J. (2012). *Technology-enhanced formative assessment in mathematics for English language learners* (Unpublished doctoral dissertation). University of Minnesota, Minneapolis. ERIC Number: ED551876.

Nunnery, J. A., & Ross, S. M. (2007). The effects of the School Renaissance program on student achievement in reading and mathematics. *Research in the Schools, 14*(1), 40–59.

Randel, B., Apthorp, H., Beesley, A. D., Clark, T. F., & Wang, X. (2016). Impacts of professional development in classroom assessment on teacher and student outcomes. *Journal of Educational Research, 109*(5), 491-502.

Renaissance Learning (2013). The research foundation for Star Assessments. Wisconsin Rapids, WI: Renaissance Learning.

Rudd, P., & Wade, P. (2006). Evaluation of Renaissance Learning mathematics and reading programs in UK Specialist and feeder schools. Slough: National Foundation for Educational Research.

Schulte, A. C., Easton, J. E., & Parker, J. (2009). Advances in treatment integrity research: Multidisciplinary perspectives on the conceptualization, measurement, and enhancement of treatment integrity. *School Psychology Review*, 38, 460–475.

Sawilowsky, S (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, 8(2), 467–474. http://digitalcommons.wayne.edu/jmasm/vol8/iss2/26/

Shang, Y., VanIwaarden, A., & Betebenner, D. W. (2015). Covariate measurement error correction for Student Growth Percentiles using the SIMEX method. *Educational Measurement: Issues and Practice, 34*(1), 4-14.

Spicuzza, R., Ysseldyke, J., Lemkuil, A., Kosciolek, S., Boys, C., & Teelucksingh, E. (2001). Effects of curriculum-based monitoring on classroom instruction and math achievement. *Journal of School Psychology, 39*(6), 521–542.

Springer, R. M., Pugalee, D., & Algozzine, B. (2007). Improving mathematics skills of high school students. *The Clearing House, 81*(1), 37–44.

Stanley, A. M. (2011). *Accelerated Mathematics and high-ability students' math achievement in grades three and four.* ProQuest LLC, Ed.D. Dissertation, East Tennessee State University. ERIC Number: ED532232.

Teelucksingh, E., Ysseldyke, J., Spicuzza, R., & Ginsburg-Block, M. (2001). *Enhancing the learning of English language learners: Consultation and a curriculum-based monitoring system.* Minneapolis, MN: University of Minnesota, National Center for Educational Outcomes.

Topping, K. J. (2017). Implementation fidelity in computerised assessment of book reading. *Computers and Education, 116*, 176-190. doi: 0.1016/j.compedu.2017.09.009.

Walker Driesel, D. (2013). *Mathematics interventions: A correlational study of the relationship between level of implementation of the Accelerated Math program and student achievement*. ProQuest LLC, Ed.D. Dissertation, Liberty University. ERIC Number: ED564886.

What Works Clearinghouse (2017). *Accelerated Math. Primary Mathematics*. Washington, DC: Institute of Education Sciences.

Wolfe, C. B., Clements, D. H., Sarama, J., & Spitler, M. E. (2013). *Sustainability of fidelity of implementation over time in the context of a prekindergarten mathematics curriculum and professional development scale-up intervention.* Washington, DC: Society for Research on Educational Effectiveness. ERIC Number: ED564101. Available https://files.eric.ed.gov/fulltext/ED564101.pdf [12 June 2018].

Wright, S. P. (2010). *An investigation of two nonparametric regression models for value-added assessment in education*. Cary, NC: SAS Institute Inc. Retrieved from https://education.ohio.gov/getattachment/Topics/Data/Report-Card-Resources/Ohio-Report-Cards/Value-Added-Technical-Reports-1/An-Investigation-of-Two-

Nonparametric-Regression-Models-for-Value-Added-Assessment-in-Education-S-Paul-Wright-1.pdf.aspx [16 June 2016].

Ysseldyke, J., & Betts, J. (2010). *Progress monitoring, implementation integrity, and guided practice across multiple math curricula.* Presentation at Council for Exceptional Children 2010 Annual Convention & Expo, Nashville, TN.

Ysseldyke, J., Betts, J., Thill, T., & Hannigan, E. (2004). Use of an instructional management system to improve mathematics skills for students in Title I programs. *Preventing School Failure, 48*(4), 10–14.

Ysseldyke, J., & Bolt, D. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review*, *36*(3), 453–467.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., & Boys, C. (2003). Effects of a learning information system on mathematics achievement and classroom structure. *Journal of Educational Research, 96*(3), 163–173.

Ysseldyke, J., Spicuzza, R., Kosciolek, S., Teelucksingh, E., Boys, C., & Lemkuil, A. (2003). Using a curriculum-based instructional management system to enhance math achievement in urban schools. *Journal of Education for Students Placed at Risk*, *8*(2), 247–265.

Ysseldyke, J., & Tardrew, S. (2007). Use of a progress-monitoring system to enable teachers to differentiate mathematics instruction. *Journal of Applied School Psychology*, *24*(1), 1–28.

Ysseldyke, J., Tardrew, S., Betts, J., Thill, T., & Hannigan, E. (2004). Use of an instructional management system to enhance math instruction of gifted and talented students. *Journal for the Education of the Gifted*, *27*(4), 293-319.

Zumwalt, D. B. (2001). *The effectiveness of computer-aided instruction in eighth-grade pre-algebra classrooms in Idaho* (Unpublished PhD dissertation). Idaho State University, Idaho. Retrieved June 12, 2018 from https://www.learntechlib.org/p/117250/.