



## **A Randomized Controlled Trial of the Morningside Math Facts Curriculum on Fluency, Stability, Endurance and Application Outcomes**

Title	A Randomized Controlled Trial of the Morningside Math Facts Curriculum on Fluency, Stability, Endurance and Application Outcomes
Author(s)	McTiernan, Aoife;Holloway, Jennifer;Healy, Olive;Hogan, Michael
Publication Date	2015-05-23
Publisher	Springer

Running Head: BUILDING FLUENT PERFORMANCES WITH MATHEMATICS

A Randomized-Controlled Trial of the Morningside Math Facts Curriculum on  
Fluency, Stability, Endurance and Application Outcomes

Mc Tiernan, A., Holloway, J., Healy, O., & Hogan, M. (in press). A Randomized Controlled Trial of the Morningside Math Facts Curriculum on Fluency, Stability, Endurance and Application Outcomes. *Journal of Behavioral Education*.

### Abstract

A randomised-controlled trial was used to evaluate the impact of a frequency-building curriculum to increase the fluency of component mathematics skills in a sample of 28 males aged 9-11 years. Assessments of mathematical ability were conducted before and after the training period to evaluate the impact of learning component skills fluently on endurance, stability and application of mathematical skills. Statistically significant differences between the experimental training group and treatment as usual control group were found on measures of fluency, endurance, stability and one subtest of the Wechsler Individual Achievement test of mathematical ability. Results indicate the efficacy of the frequency-building curriculum in promoting fluency with component skills. Results are discussed in light of research and theory in the area of instructional design and behavioral fluency enhancement.

*Keywords:* Mathematics; frequency-building; fluency outcomes; randomised-controlled trial

Mathematics is an important component of the educational curriculum. Competency in basic mathematics is not only necessary for completing related classwork but is also essential for adaptive functioning in everyday life. Geary, Hoard, Nugent, and Bailey (2012) equate the importance of mathematical knowledge with that of basic literacy and emphasise its significance for independent functioning in society. The development of mathematical expertise facilitates advancement in numerous other areas including business, medicine, and engineering; and how students later perform mathematically also impacts labour quality and national growth (Lin & Kubina, 2005).

The findings from the 2009 Program for International Assessment demonstrated that students from the USA performed significantly below the Organization for Economic Co-operation and Development (OECD) average in mathematics, with students from Ireland similarly performing below average (OECD, 2010). Since 2009, there has been a relative improvement in the performance of Irish students to just above the OECD average, while students in USA continued to perform below average (OECD, 2014). Low performances in mathematics have prompted investigations into instructional programs and interventions which could enhance students' development of mathematical skills (Coddington et al. 2010; Poncy et al. 2013).

Proponents of fluency and generative instruction argue that fluency with basic mathematics skills is pivotal to success in progressing adeptly to more complex learning in the area (Johnson & Street, 2013). Generative instruction involves breaking down each teaching goal into sets of tool skills, component skills and composite skills (Johnson & Street, 2013). Tool skills are the basic pre-requisites that are necessary to perform a number of more complex skills, for example, holding a

pencil, writing marks on a page with a pencil. Johnson and Street (2013) describe component skills as second-level building blocks which depend upon tool skills but are pre-requisite skills necessary for engaging in composite behaviors, for example, writing numbers and writing letters. Composite skills are the higher level performances that represent mastery of a content area and comprise combinations of component skills, for example, developing solutions and writing answers to mathematics problems (Johnson & Street, 2013). Johnson and Layng (1994) maintain that the combination of component-composite analysis of behavior and cumulative instruction using procedures that increase fluency with component skills are fundamental for an effective system of generative instruction.

Consistent with this theory, a link between fluent performances of component mathematics skills and overall mathematical ability has been reported in the literature. In a sample of 241 students in the second grade, Carr, Steiner, Kyser, and Biddlecomb (2008) found that fluency with single-digit arithmetic problems (i.e., addition and subtraction problems) significantly predicted mathematical ability which was measured using a Criterion Referenced Competency Test (CRCT). In a follow up study, Carr and Alexeev (2011) examined growth trajectories in mathematical ability with 240 participants who had taken part in the previous study. While the findings in the 2008 study demonstrated that fluency, accuracy and gender all influenced mathematical ability, the results of the 2011 study found that fluency had the most significant impact on growth in mathematical ability. Such findings emphasise the importance of building fluency with component skills. However, research to date has not examined the impact of fluency-based instruction with component skills on overall mathematical ability as measured by standardised assessments of achievement.

Early findings investigating behavioral fluency in educational settings, and more recent emerging research in the area, suggest that additional outcomes, critical to learning and education, are associated with fluent performances of component academic skills. Haughton (1980) first identified critical outcomes associated with fluency using the acronym REA-PS which represented Retention, Endurance, and Application Performance Standards. The most recent acronym to describe outcomes that are associated with performance fluency is MESAG-PS (Johnson & Street, 2013) which represents Maintenance, Endurance, Stability, Application and Generativity Performance Standards.

*Endurance* involves the ability to engage in a skill for prolonged periods without fatiguing (Fabrizio & Moors, 2003) or the ability to meet real-world requirements for how long the behavior should be performed (Johnson & Street, 2013). Binder, Haughton, and Van Eyk (1990) equate endurance to the ability to maintain “attention” to a task. Johnson and Layng (1992) further separated endurance into two components, performance over extended periods of time (endurance) and performance in the presence of distracting stimuli (stability). *Stability* is described as the ability to continue to perform a skill amidst distractions (Johnson & Street, 2013). Despite the importance of such skills within educational settings, the ability to attain these outcomes as a consequence of fluency-based instruction has only been demonstrated in a small number of research studies (Brady & Kubina, 2010; McDowell & Keenan, 2001).

Kubina and Yurich (2012) describe *application* as the ability to apply one or more element behaviors or pre-requisite skills to more complex behaviors or skills once a specific frequency with those skills is achieved (e.g., the ability to recognise or read letters at a high frequency will result in an improvement in applying this skill to

decoding full words). A number of studies to date have demonstrated the ability of fluency-based instruction to improve participants' application of fluent skills to more complex tasks (Bucklin, Dickinson, & Brethower, 2000; Cavallini & Perini, 2009; Chiesa & Roberston, 2000; Kubina, Young, & Kilwein, 2004). However, continued research is necessary in order to further evaluate the occurrence of this outcome as a consequence of fluency-based instruction across behavioral repertoires.

Despite such positive outcomes associated with fluent performances, research is necessary to identify the most effective, evidence-based interventions and curricula to do so. Coddling, Hilt-Panahon, Panahon, and Benson (2009) conducted a meta-analysis to investigate mathematics interventions with students identified as needing additional support with mathematics. Single-subject research design (SSRD) was used in 68% of the studies, with a limited number of studies employed to evaluate each intervention. The majority of interventions evaluated did not meet criteria for evidence-based practice. Such findings indicate a necessity for continued empirical evaluation of fluency-based instruction and the necessity to conduct research with group based designs. The research area would benefit from the evaluation of fluency-based instruction using randomised controlled trials.

Frequency-building is a procedure which has been demonstrated as a promising approach to attaining fluency and has been described as the timed repetition of selected behavior followed by performance feedback (Kubina & Yurich, 2012). Timed repetition of skills or "frequency-building" can be incorporated into mathematics instruction to increase rate of correct responding with component mathematics skills. Often within behavioral fluency research and practices, Explicit Timing (ET) is used to allow timed practice of component skills and achieve fluent performances. ET involves the presentation of a task and a specific amount of time

allocated to complete it (Gross et al., 2013). Timings, which are usually one minute in duration, are conducted during which students complete as many repetitions of the target skills as they can before the timing ends. Research to date has demonstrated promising outcomes in relation to such frequency-building strategies to increase mathematics fluency (Hartnedy, Mozzoni, & Fahoum, 2005; Poncy, Duhon, Lee, & Key, 2010).

The generative instruction approach to teaching academic skills involves cumulative, fluency-based instruction with skills that have been identified as key components (Johnson & Layng, 1994). Based on this approach, a number of instructional curricula have been developed to teach academic skills in this manner. Specifically, the Morningside Mathematics Fluency: Math Facts (Johnson, 2008) curricula are designed to teach key component mathematics skills, namely addition, subtraction, multiplication and division. The curricula are used in conjunction with frequency-building and are designed to facilitate the achievement of fluent performances with such skills. Johnson and Street (2013) describe informal evaluations of their use, in conjunction with other curricula, across three schools in the U.S., resulting in increases in participants' scores on standardised assessments of mathematical ability. Empirical evaluation of the Morningside Mathematics Fluency: Math Facts curriculum (Johnson, 2008) would be beneficial in order to add to the literature examining effective, evidence-based instructional approaches for teaching mathematics skills.

### **Purpose of the Present Study**

The current research employed a randomised controlled trial (RCT) to evaluate the effects of frequency-building using the Morningside Mathematics Fluency: Math Facts (Johnson, 2008) in comparison to a treatment as usual control



(TAU) condition in a sample of 28 males aged 9-11 years. A number of research studies to date have demonstrated the attainment of fluent performances as a result of frequency-building with mathematics skills (Hartnedy et al., 2005; Poncy, et al., 2010). Further, the attainment of fluent performances with component mathematics skills has been associated with improvements in endurance (Brady & Kubina, 2010; McDowell & Keenan, 2001) and application (Bucklin et al., 2000; Cavallini & Perini, 2009; Chiesa & Roberston, 2000; Kubina et al., 2004) as well as overall mathematical ability (Carr & Alexeev, 2011; Carr et al., 2008). For this reason, pre- and post-test measures of fluency with targeted math skills, endurance, stability and application, and the Wechsler Individual Achievement Test Second Edition (WIAT-II; Wechsler, 2005) mathematics subtests were recorded for both groups. Controlling for baseline differences, it was hypothesised that the frequency-building group would show significantly greater performance than the Control Group for all post-intervention outcome measures.

## **Method**

### **Participants and Setting**

Thirty-six males attending an all-boys primary school in Ireland participated in the study. The participating school was recruited through convenience sampling after which the teachers were asked to identify students from fourth and fifth grade performing below age equivalency in mathematics. All participants were identified as demonstrating difficulties in performing at an age equivalent level on the mainstream mathematics curriculum and were receiving daily remedial support in mathematics. Subsequently, participants were assessed using the WIAT-II (Wechsler, 2005) and were included only if they scored below 100. For this reason, one participant was excluded from the study having achieved a composite mathematics standard score of

108 at pre-testing. Participants were matched with a peer whose standard score on the WIAT-II (Wechsler, 2005) mathematics subtests was no more than ten points more or less than their own. Consequently, only 28 of the remaining 35 participants could be matched, which excluded a further seven participants from the study. The final sample consisted of 28 males with a mean age of 10 years 1 month ( $SD = .64$ , range: 9-11.4). Participants' mean composite score on the Mathematics sub-tests of the WIAT-II (Wechsler, 2005) was 86.6 ( $SD = 9.23$ , range: 61-99).

The intervention took place during school hours with 3-4 participants and one instructor in each session. Sessions were conducted with fourth and fifth grade groups separately. Within each grade, participants were randomly assigned to either the Experimental Group or the Control Group condition. Sessions were conducted in a small classroom with all participants sitting at a semi-circular table.

### **Intervention agents and training**

The instructor was a Masters level student completing university postgraduate training in Applied Behavior Analysis. Training sessions in frequency-building and use of the curriculum were provided by the experimenter prior to and throughout the course of the school year. The experimenter was on-site each week to observe between one and two sessions and provided feedback to the student on the implementation of the intervention.

### **Materials**

Each participant receiving intervention was allocated a folder with materials for frequency-building. Materials included a sticker chart to self-monitor and reinforce achievement of fluency aims. The chart consisted of three columns. The first column listed each worksheet in sequence and the second column contained the aim for each worksheet. When the aim for each worksheet was achieved,

participants received a sticker which they placed in the third column next to their aim. The folders also included the worksheets to be completed within the session. Digital timers were used to conduct timings.

**Curriculum.** The Morningside Math Facts: Multiplication and Division curriculum (Johnson, 2008) was used to teach component skills in mathematics. This curriculum is designed to build fluency with multiplication and division computation. Fact families are used to build fluency with simple multiplication and division facts. This method is purported to reduce the instructional time necessary to master such skills. Once a student learns one number family (e.g., 2, 3, 6) it yields four separate math facts (i.e.,  $2 \times 3 = 6$ ,  $3 \times 2 = 6$ ,  $6 \div 2 = 3$ ,  $6 \div 3 = 2$ ). A worksheet containing target fact families is provided so that each fact family can be recited both accurately and fluently. The curriculum includes a fluency aim of accurately reciting each fact family within 4-6 seconds before progressing to pencil and paper worksheets.

Pencil and paper worksheets each consist of 100 problems presented in random order pertaining to 36 fact families and multiplication and division by one and zero. The fluency aim for number of correct responses per minute for each worksheet is 50-60 per minute. The curriculum also includes cumulative and review worksheets which consist of problems pertaining to fact families learned in previous sessions. The aims set by the curriculum are 60-70 correct responses per minute for cumulative worksheets and 70-80 correct responses per minute for review worksheets. All problems require one or two digit answers. A correct response is recorded if all digits in the answer to the problem are correct. Responses are scored as incorrect if any digits in the answer to a problem are incorrect, if digits are omitted or placed in the incorrect order. There are a total of 84 fluency aims to achieve throughout the complete curriculum.

**Dependent Measures**

Dependent variables included the WIAT-II (Wechsler, 2005), fluency with targeted mathematics skills and three of the fluency outcomes defined by Binder (1996), namely, endurance, stability and application.

**WIAT-II.** Participants were assessed using the WIAT-II (Wechsler, 2005) as a standardised measure of mathematics ability. The WIAT-II (Wechsler, 2005) is a norm-referenced, individually administered test of academic achievement. Internal consistency ranges from .80-.98 and test-retest reliability ranges from .85-.98 (Wechsler, 2005). Both Numerical Operations and Mathematical Reasoning subtests were administered. The Numerical Operations subtest is a pencil and paper test with items that progress in complexity assessing numerical identification, counting, numeral writing, simple addition and subtraction, integer arithmetic, multi-digit addition and subtraction, single digit multiplication and division, fractions, decimals and percentages as well as integers. Mathematical Reasoning is an orally presented verbal problem-solving test with pictures. This test assesses counting, comparison, simple addition and subtraction word problems, mathematical language, interpretation of charts, completion of patterns, knowledge of measures and money, fractions, decimals, probability and mental rotation.

**Fluency.** A review worksheet from the Morningside Math Facts: Multiplication and Division curriculum (Johnson, 2008) was used to assess fluency with mathematics skills targeted during intervention. Each participant was assessed using this worksheet which provided multiple exemplars of multiplication and division problems which were targeted throughout the frequency-building intervention. One hundred problems relevant to all 12 fact families were presented in random order on the worksheet. Participants completed as many problems as they

could during a 1-minute timing. Rate of correct responding per minute was calculated by the experimenter. All math problems required one or two digit answers. A correct response was recorded if all digits in the answer to the problem were correct.

Responses were scored as incorrect if any digits in the answer to a problem were incorrect, if digits were omitted or placed in the incorrect order.

**Endurance.** In order to evaluate the ability to sustain a steady performance over an increased length of time, participants completed as many problems as they could, using pencil and paper worksheets during a 5-minute timing. Four worksheets containing problems that were presented during the intervention with the Experimental Group were presented with one hundred problems on each. Measures of both rate of correct responding (Brady & Kubina, 2010) and on-task behavior (McDowell & Keenan, 2001) were recorded.

**Rate of correct responding.** Rate of correct responding per minute was calculated to assess if participants could sustain a high rate of correct responding during prolonged timings to which they were not exposed during frequency-building sessions. As with the measure of fluency, a correct response was recorded if all digits in the answer to the problem were correct. Responses were scored as incorrect if any digits in the answer to a problem were incorrect, if digits were omitted or placed in the incorrect order. Rate of correct responding per minute was calculated by dividing the number of correct responses by five to attain participants' average rate of responding per minute.

**Systematic observation of on-task behavior.** Each participant was also systematically observed during the same 5-minute timing using the Behavioral Observation of Students in Schools (BOSS; Shapiro, 2004). The BOSS (Shapiro, 2004) is a measure that is used to directly observe active engagement on academic

tasks (Chafouleas, Sanetti, Kilgus, & Maggin, 2012). Proponents of fluency suggest that achievement of high performance frequencies increases individuals' ability to maintain attention to a task over extended durations of performance. For this reason, operational definitions from the BOSS (Shapiro, 2004) were used to observe active engagement in tasks during extended timings to assess this critical learning outcome.

The BOSS (Shapiro, 2004) divides on-task behavior into active engaged time (AET) and passive engaged time (PET). Active engaged time is defined as when a student is actively engaged in academic responding (e.g., reading aloud or writing). Passive engaged time is defined as when a student is passively attending (e.g., listening during instruction). As participants should be actively engaged (i.e., writing answers to multiplication and division problems) when completing written mathematics problems, only AET was recorded. Momentary-time sampling with ten second intervals was used to measure the percentage of intervals participants were observed to be actively engaged.

**Stability.** In order to evaluate the ability to sustain a steady performance over an increased length of time in the presence of distractors, 3-minute timings were conducted during which participants completed pencil and paper worksheets with problems presented during the intervention phase. However, during each participant's stability timing, two or three students were instructed to complete oral math problems aloud within a small group next to them. This was implemented in an effort to create a distraction for the target participant taking the stability test. Similar to measures of endurance, the worksheets were scored by the experimenter and rate of correct responding per minute was calculated. Rate of correct responding per minute was calculated by dividing the number of correct responses by three to attain participants' average rate of responding per minute. On-task behavior was also observed during

stability timings using momentary-time sampling and the BOSS (Shapiro, 2004). The percentage of intervals participants were observed to be actively engaged was again recorded.

**Application.** To assess participants' ability to apply component skills targeted during intervention to more complex mathematics problems, a test consisting of more complex single and double digit multiplication and division problems was administered. Problems included in the test were taken from the school curriculum math book used for the previous academic year. All of the problems relied on division and multiplication of digits within fact families targeted during the fluency intervention. Participants completed as many problems as they could on the test during a 2-minute timing. The test was scored by the experimenter and digits correct per minute (DCM) was calculated. Each correct digit within an answer was scored as correct if it was placed in the correct position in the answer. Digits were scored as incorrect if an incorrect digit was written, if it was incorrectly placed in the answer or if it was omitted

### **Independent Measures**

**Control condition.** Participants in the Control Group participated in TAU which included typical classroom instruction and additional learning support classes. Learning support classes were conducted three times per week in small groups of four to five for 30 minutes. These classes consisted of additional supplemental instruction in mathematics which was designed by the learning support teacher and based on individual student needs in line with recommendations from the Department of Education and Science (DES, 2000). The teacher was a qualified primary school teacher with a Bachelor of Education. Participants in the Control Group did not receive instruction with the Morningside Curriculum nor did they take part in

frequency-building with any instructional materials during typical classroom instruction or learning support classes.

**Experimental condition.** Participants in the Experimental Group received TAU identical to that received by the Control Group. In addition to TAU, experimental participants were exposed to the frequency-building intervention for twenty minute sessions, one day per week until they had achieved 26 fluency aims targeted for intervention with the Experimental Group only. The 26 aims pertained to the first 12 fact families in the curriculum. Fluency aims (i.e., target number of correct responses per minute) were pre-determined by the curriculum for each worksheet. During each frequency-building session, ET was conducted with participants while they completed worksheets from the curriculum. Participants moved to the next worksheet, in the sequence of the curriculum, when they had achieved their aim on the previous. Therefore, the length of time spent on learning targets orally and on pencil and paper worksheets varied according to individual participants' performance on each. As each experimental participant achieved all targeted 26 aims, post-tests were conducted with that participant and their matched control.

### **Procedure**

**Curriculum start point allocation.** Prior to implementing the intervention, one-minute timed probes were conducted with the first two targets in the curriculum to investigate whether participants in the Experimental Group could achieve fluency targets as outlined by the Morningside curriculum. This was conducted to ensure that no participant would start at a point in the curriculum at which they were already fluent. No participant reached fluency aims on timed probes and all were allocated a start point at the beginning of the curriculum.



**Frequency-building sessions.** During the first session, folders containing the instructional materials were allocated and the general procedure was explained to the group. All participants were told that they would receive a sticker each time they achieved a fluency aim and that their terminal goal was to achieve 26 fluency aims. During subsequent sessions, participants were instructed to open their folders and find the worksheet they were currently working on using their sticker charts as a reference. The pre-determined aims were also identified by participants using the chart.

Before participants began 1-minute timings on each worksheet, they were required to learn each fact family orally. It was explained that they should practice reciting each math fact as fast as they could. The participants were also informed that their aim was to recite each fact family in 4-6 seconds. Each participant practiced this independently while the remaining participants engaged in either ET with paper and pencil worksheets or practiced learning their own targets orally. The instructor then tested each participant to verify that they had achieved their aim of reciting each fact family in 4-6 seconds.

When participants were working on pencil and paper worksheets, they identified their fluency aim using their sticker chart and marked an X on their worksheet as a visual prompt to indicate how many problems should be completed correctly to achieve that aim. The instructor monitored participants to ensure that everyone in the group was working on the correct target and had identified the correct aim before starting. When the group were ready, the timing began. Each participant was instructed to place their pencil at the start point on their worksheet. They were reminded that they needed to answer as many problems correctly as they could within one minute and should aim to beat their score from the last timing. The instructor set the timer for one minute, said “Let’s begin” and started the digital

timer. All of the participants within the small group session completed their own worksheets at the same time.

**Corrective feedback.** No corrective feedback was delivered during timings; however, participants received praise throughout the timing contingent only on active engagement in the task. Once a timing was complete, the group were instructed to “drop their pencils”, count how many problems they had completed and to signify any incorrect responses by drawing a circle around it. The instructor monitored the group to verify number of correct and incorrect responses and to provide corrective feedback on errors with each participant individually. The instructor identified incorrect responses and presented the relevant problems orally. Least-to-most prompts were used until the participant could emit the correct response (e.g., the problem was re-presented orally to the participant, the participant was asked to recite the relevant fact family to help solve the problem, the participant was asked to refer to their written version of the relevant fact family to solve the problem, the instructor modelled the correct response for the participant.)

Praise, the opportunity to choose a sticker, and progression to the next worksheet were contingent on the attainment of each fluency aim. When a fluency aim was not attained, additional timings were subsequently completed on that worksheet. One-minute timings were used for all paper and pencil worksheets; therefore, participants did not have to work on the same targets as their peers in the group. Each could progress to their next worksheet as soon as they achieved their aim. Participants received praise for higher rates of responding and for correct responding e.g. “You beat your score from the last timing, well done” or “You answered all of the problems correctly, well done”.

### **Inter-Rater Reliability**

**Rate of correct responding.** Thirty percent of the worksheets were independently scored by two individuals to examine inter-rater reliability. Inter-rater agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplied by one (Coddling et al., 2010). Percentage agreement for number of correct responses on worksheets was 100%.

**Observation of on-task behavior.** Inter-observer agreement was conducted for 29% of observations using the BOSS (Shapiro, 2004). Two observers, the experimenter and a trained independent observer, independently collected inter-observer agreement data during both endurance and stability timings. The independent observer was blind to the group to which each participant had been assigned. Each observer rated whether the participant was actively engaged or off-task using momentary-time sampling with ten second intervals. The end of each interval was signalled using an audio cue which both observers listened to through headphones. Inter-observer agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplied by one hundred. Percentage agreement for number of the number of intervals observed on- or off-task was 98.8% (Range = 83% - 100%).

### **Social Validity**

A brief questionnaire to assess social validity was designed by the experimenter and administered subsequent to the intervention (see Appendix). Eleven of the fourteen participants in the Experimental Group completed the questionnaire. The remaining three participants were absent during the administration of this questionnaire, which was followed directly by school holidays. Therefore, they were not available to complete this measure. Questions focused on the participants' enjoyment of fluency training, most difficult and easiest elements of the intervention

and whether or not they thought it had helped them. Participants were required to select “yes” or “no” after reading a question related to the intervention.

### **Design and Analysis**

A parallel group Randomized Control Trial (RCT) was used to evaluate the efficacy of the frequency-building intervention. Participants across groups were matched according to current school grade, composite mathematics scores on the WIAT-II (Wechsler, 2005) and age, and were subsequently randomly assigned to either an experimental ( $n = 14$ ) or control ( $n = 14$ ) condition. Participants were matched with a peer in the same grade (dyads) and with those whose standard score on the WIAT-II (Wechsler, 2005) mathematics subtests was no more than ten points more or less than their own. After meeting these criteria, participants closest in age were then matched.

Dependent measures were recorded for all participants. The Experimental Group received the frequency-building intervention while both groups participated in classroom instruction and learning support classes as usual (i.e., the Control Group represented a TAU group). Once a participant in the Experimental Group had completed the intervention, dependent measures were administered again with that participant and their matched control. The experimenter conducted pre- and post-test assessments with all participants while a trained instructor carried out the frequency-building intervention with the Experimental Group. The intervention was run over 32 weeks, inclusive of pre- and post-testing, with a mean of 23.4 ( $SD = 4.01$ ) sessions implemented with participants in the Experimental Group.

A between-groups multivariate analysis of covariance (MANCOVA) was conducted to investigate differences between the Experimental and Control Group on all 8 dependent variables at post-testing. Participants’ pre-test scores were used as the

covariates in the analyses. Preliminary checks were conducted for each MANCOVA to ensure that there was no violation of assumptions. Post-hoc analyses were conducted for each variable using paired samples t-tests to investigate statistical differences in scores from pre- to post-testing for both the Experimental and Control Group.

### **Results**

The Experimental Group progressed through the frequency-building intervention with the majority (71%) of participants achieving all 26 fluency aims. The remaining 29% achieved between 13 and 24 aims. Fluency aims varied from 50-70 correct responses per minute across the target worksheets. Table 1 provides information on the number of aims achieved and number of sessions completed with each participant.

MANCOVA revealed statistically significant difference between groups on the combined dependent variables,  $F(8, 11) = 13.3, p < .001; \lambda = .094; \eta_p^2 = .906$ . When considered separately, statistical differences between groups on post-test scores were found for the Mathematical Reasoning subtest of the WIAT-II (Wechsler, 2005) and for rates of correct responding on tests of fluency, endurance and stability. No significant differences were observed for post-test scores on the Numerical Operations subtest of the WIAT-II (Wechsler, 2005), BOSS (Shapiro, 2004) scores, or rates of correct responding on tests of application. These results are described in more detail below. Table 2 summarises the mean performance for both groups on each dependent measure at both pre- and post-tests.

#### **WIAT-II**

A significant difference was found between the Experimental and Control Group on the Mathematical Reasoning sub-test,  $F(1, 18) = 4.55, p = .047, \eta_p^2 = .202$ .

The Experimental Group scored higher than the Control Group at post-test. Paired samples t-tests also revealed a significant increase in the Experimental Group's scores on the Mathematical Reasoning subtest from pre- to post-test,  $t(13) = 2.52, p = .025, \eta^2 = .333$ . Conversely, no significant increase on the Mathematical Reasoning subtest was observed for the Control Group,  $t(13) = .211, p = .836$ . No significant differences between groups were observed for the Numerical Operations score nor were there significant increases in scores on this measures from pre-test to post-test for either group.

### **Fluency**

There was a significant difference between groups on post-test scores for targeted mathematics skills,  $F(1, 18) = 24.5, p < .001, \eta_p^2 = .577$ , with the Experimental Group's rate of responding being significantly higher than that of the Control Group. Paired samples t-tests also showed a statistically significant increase for the Experimental Group in rate of correct responding from pre-test to post-test,  $t(13) = 7.73, p < .001, \eta^2 = .832$ . There was also a significant increase for the Control Group,  $t(13) = 3.43, p = .004, \eta^2 = .495$ , suggesting improvements associated with practice or standard remediation work in class.

### **Endurance**

There was a significant difference between the two groups on post-test endurance rates of correct responding,  $F(1, 18) = 86.1, p < .001, \eta_p^2 = .827$ . The Experimental Group demonstrated a statistically significant increase in rate of correct responding during endurance timings from pre- to post-tests,  $t(13) = 11.9, p < .001, \eta^2 = .922$ . A significant increase in rate of correct responding was also found for the Control Group,  $t(13) = 4.56, p = .001, \eta^2 = .613$ .

No significant difference was found between groups on percentage of intervals observed on-task,  $F(1, 18) = .736, p = .402, \eta_p^2 = .039$ . Both groups were observed to stay on-task for a higher percentage of intervals at post-tests. The mean percentage of intervals observed on-task increased significantly from pre-to post-test for the Experimental Group,  $t(13) = 4.10, p = .001, \eta^2 = .583$ , and the Control Group,  $t(13) = 3.6, p = .003, \eta^2 = .519$ .

### **Stability**

There was a significant difference between groups on post-test rates of correct responding during stability timings,  $F(1, 18) = 53.5, p < .001, \eta_p^2 = .748$ . Paired samples t-tests showed a statistically significant increase in rate of correct responding for the Experimental Group from pre- to post-testing,  $t(13) = 13.6, p < .001, \eta^2 = .939$  and for the Control Group,  $t(13) = 6.04, p < .001, \eta^2 = .752$ . Thus, while both groups demonstrated significant improvements, those who had been exposed to frequency-building were able to maintain a higher rate of correct responding in the presence of a distractor.

Percentage of intervals observed on-task increased across all participants during stability timings with no significant differences between groups at post-tests,  $F(1, 18) = .063, p = .805, \eta_p^2 = .003$ . The mean percentage of intervals observed on-task increased significantly from pre- to post-testing for the Experimental Group,  $t(13) = 3.19, p = .008, \eta^2 = .449$ , and for the Control Group,  $t(13) = 3.75, p = .002, \eta^2 = .54$ .

### **Application**

No significant difference between groups was found for application performance at post-test,  $F(1, 18) = 1.26, p = .276, \eta_p^2 = .065$ . Paired samples t-tests

for application timings showed a statistically significant increase in rate of correct responding for the Experimental Group from pre- to post-tests,  $t(13) = 7.94, p < .001, \eta^2 = .84$ , and for the Control Group from pre- to post-test,  $t(13) = 7.2, p < .001, \eta^2 = .812$ .

### **Social Validity**

Eighty-one percent of participants within the Experimental Group responded that they would like to learn more mathematics problems fluently; 100% of participants responded that they learned their math facts well, remembered them better after the intervention and liked achieving their aims; 90% responded that they liked practicing their mathematics as fast as they could with a timer and 82% responded that they were quicker with their maths. Forty-five percent of participants responded that they found it difficult to achieve their aims while 55% responded that they did not find it difficult.

### **Discussion**

Using an RCT, the current research demonstrated positive outcomes associated with a frequency-building intervention using the Morningside Math Facts: Multiplication and Division curriculum (Johnson, 2008). Children between the ages of 9 and 11 years, identified as demonstrating difficulties performing at an age equivalent level on the mainstream mathematics curriculum, demonstrated increases in rate of correct responding with targeted component mathematics skills. The majority of participants completed the intervention curriculum, demonstrated enhanced performance on critical fluency outcomes and also demonstrated transfer of benefits to standardised tests of mathematical ability.

Researchers suggest that promoting accuracy and fluency with mathematics skills should be a focus of instruction in primary school years (Carr & Alexeev,



2011); however, a gap in the research exists with respect to empirical validation of fluency-based instruction (Coddington et al., 2009; Poncy et al., 2013). The findings of the current research contribute to the literature with important implications for applied settings. The results are in accordance with previous research showing the efficacy of frequency-building to increase rates of correct responding in mathematics skills with primary school-aged children (Chiesa & Robertson, 2000; Gross et al., 2013; Hartnedy et al., 2005; Poncy et al., 2010). Further, while continued research should be conducted to further validate the Morningside Math Facts: Multiplication and Division curriculum (Johnson, 2008), the current findings provide evidence for its efficacy and provide sufficient rationale for its incorporation into educational settings to increase fluency with component mathematics skills.

To date, few studies have investigated the ability of such interventions to impact standardised scores of mathematical ability. The current study found increases in scores on the Mathematical Reasoning subtest of the WIAT-II (Wechsler, 2005) only for the Experimental Group. There was no change in score for the Control Group. Although results indicated a small effect size, frequency-building implemented only once per week for 15-20 minutes across 32 weeks, had a significant impact on standardised measure of Mathematical Reasoning. This finding emphasises the value of teaching component skills fluently with respect to generalised effects as measured by the WIAT-II (Wechsler, 2005).

Critical learning outcomes investigated in the current study included endurance, stability and application. Research investigating critical outcomes to date focuses primarily on application and retention with fewer examinations of endurance reported in the literature (Brady & Kubina, 2010). To date, no empirical investigations of this outcome have been reported in the literature. Findings from the

current study represent a significant contribution to the literature in this regard. The attainment of fluent performances resulted in improvements in endurance and stability and is consistent with the behavioral fluency literature to date. Further, measures of performance (i.e., rate of correct responding) were found to be more sensitive to the effects of building fluency than systematic observation of on-task behavior using the BOSS (Shapiro, 2004) thus adding to the literature on how best to measure these outcomes.

Relative to endurance and stability, application has been investigated to a greater extent. However, additional research is warranted to provide evidence of improvements in this outcome once fluent performances are attained. In the current study, no significant difference between groups was found on measures of application. Research in the area of behavioral fluency would suggest that experimental participants' scores on the test of application should have improved significantly more when compared with those who have not achieved fluency with component skills (Binder, 1996; Chiesa & Robertson, 2000). The automaticity of these skills should allow them to be more readily applied.

Conversely, other researchers discuss the importance of practicing the application of fluent component skills rather than these automatically emerging once fluent (Coddling et al., 2010; Martens & Witt, 2004). It has been suggested that two types of practice exist consisting of drills and composite practice (Coddling et al., 2010; Cohen et al., 1992; Haring & Eaton, 1978). Drills involve the practice of isolated items. Composite practice requires the use of learned component responses in combination with previously learned responses. It is possible that drill and composite practice each serve separate roles when building fluency with component skills (Coddling et al., 2010; Cohen et al., 1992; Haring & Eaton, 1978) and may be

equally as important. Future research should address the effects of both types of practice and associated outcomes.

The social validity questionnaire was completed by 81% of the sample. The majority of participants reported that they enjoyed taking part in the intervention. This has important implications for the application of frequency-building with students in educational settings. It is suggested that low rates of responding when completing mathematics tasks, within set time limits, can be associated with receiving no contingent reinforcement, delayed, or less frequent reinforcement as a consequence of infrequently completing such tasks (Bliss et al., 2010; Logan & Skinner, 1998; Skinner, 2002). This, in conjunction with a high level of response effort, is purported to decrease the probability that students engage in mathematics tasks (Billington et al., 2004; Bliss et al., 2010; Skinner et al., 2005). Frequency-building incorporates positive reinforcement contingent on higher rates of correct responding following each timing which may increase students' probability to engage. Reported positive views of the intervention are consistent with this in that the majority of participants reported motivation to engage in this type of instruction in the future.

### **Limitations and Future Directions**

Participants who did not achieve the target aims on successive weeks did not meet their terminal goal due to time constraints imposed by the end of the school year. Larger gains may have been obtained on tests of generalised outcomes had each participant in the Experimental Group attained the terminal goal. Such gains may have been further impacted by the intensity of the intervention. Lipsey and Hurley (2009) maintain that the stronger the intervention "dose" that is applied during intervention, the larger the effects of the intervention. Future research should examine the impact of a higher intervention "dose" on outcome measures.

Both subtests of the WIAT-II (Wechsler, 2005) target a wide range of mathematical concepts and skills which required the use of many more component skills other than those targeted within this study. The current study targeted only 26 fluency aims from a possible 84, which represents the achievement of fluency with 12 multiplication and division fact families from a possible 36 fact families. Frequency-building interventions targeting additional component skills necessary for completing such tests (e.g., number reading and writing fluency, addition and subtraction fluency, fluency with fractions and word problems) and the achievement of fluency with all 36 fact families may demonstrate more substantial gains on such measures. Future longitudinal research should investigate the impact of frequency-building with a larger number and more diverse set of component skills on standardised measures of mathematical ability.

Training sessions were provided to the instructor throughout the course of the intervention and the experimenter was on-site weekly to observe and provide feedback. However, a limitation of the current study was the lack of treatment integrity data obtained to demonstrate the fidelity of treatment implementation by the instructor. Future studies should ensure to obtain such data in order to report on the accuracy and reliability of the independent variable.

## **Conclusion**

Competency in mathematics is pivotal for engaging in many academic and applied skills and is essential to everyday life. Proficiency in this academic domain has been linked to fluent performances of key component skills (Carr & Alexeev, 2011; Carr et al., 2008). However, there are a paucity of evidence-based interventions available to increase fluent performances with such skills (Coddington et al., 2009). The current study employed an RCT to evaluate frequency-building using

the Morningside Math Facts: Multiplication and Division curriculum (Johnson, 2008).

Results demonstrated the efficacy the intervention, adding significantly to literature examining evidence-based practices for mathematics instruction.

Positive outcomes were not only demonstrated across measures of fluency but also the attainment of fluency was found to improve performances on measures of critical learning outcomes and on standardised assessments of general mathematical ability. Coddling et al. (2009) report their results of a meta-analysis investigating fluency-based instruction with mathematics skills, noting that very few studies examine associated generalised effects. Rate of correct responding is the predominant dependent variable investigated in research studies to date. The current findings contribute considerably to the literature in this area and provide a significant rationale for the promotion of fluent component repertoires to improve general mathematical ability.

### References

- Billington, E. J., Skinner, C. H., & Cruchon, N. M. (2004). Improving sixth-grade students' perceptions of high-effort assignments by assigning more work: Interaction of additive interspersal and assignment effort on assignment choice. *Journal of School Psychology, 42*(6), 477-490. doi: 10.1016/j.jsp.2004.08.003
- Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*(2), 163.
- Binder, C., Haughton, E., & Van Eyk, D. (1990). Precision teaching attention span. *Teaching Exceptional Children, 22*(3), 24-27.
- Bliss, S. L., Skinner, C. H., McCallum, E., Saecker, L. B., Rowland-Bryant, E., & Brown, K. S. (2010). A comparison of taped problems with and without a brief post-treatment assessment on multiplication fluency. *Journal of Behavioral Education, 19*(2), 156-168. doi: 10.1007/s10864-010-9106-5
- Brady, K. K., & Kubina, R. M. (2010). Endurance of multiplication fact fluency for students with attention deficit hyperactivity disorder. *Behavior modification, 34*(2), 79-93. doi: 10.1177/0145445510361331
- Bucklin, B.R., Dickinson, A.M., & Brethower, D.M. (2000). A comparison of the effects of fluency training and accuracy training on application and retention. *Performance Improvement Quarterly, 13*, 140-163. doi: 10.1111/j.1937-8327.2000.tb00180.x
- Carr, M., & Alexeev, N. (2011). Fluency, accuracy, and gender predict developmental trajectories of arithmetic strategies. *Journal of Educational Psychology, 103*(3), 617. doi: 10.1037/a0023864

- Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*(1), 61-75. doi: 10.1016/j.lindif.2007.04.005
- Cavallini, F., & Perini, S. (2009). Comparison of teaching syllables or words on reading rate. *European Journal of Behavior Analysis, 10*, 225-263.
- Chafouleas, S. M., Sanetti, L. M., Kilgus, S. P., & Maggin, D. M. (2012). Evaluating sensitivity to behavioral change using direct behavior rating single-item scales. *Exceptional Children, 78*(4), 491-505.
- Chiesa, M., & Robertson, A. (2000). Precision teaching and fluency training: Making maths easier for pupils and teachers. *Educational Psychology in Practice, 16*(3), 297-310. doi: 10.1080/713666088
- Codding, R. S., Archer, J., & Connell, J. (2010). A systematic replication and extension of using incremental rehearsal to improve multiplication skills: An investigation of generalization. *Journal of Behavioral Education, 19*(1), 93-105. doi: 10.1007/s10864-010-9102-9
- Codding, R. S., Hilt-Panahon, A., Panahon, C. J., & Benson, J. L. (2009). Addressing mathematics computation problems: A review of simple and moderate intensity interventions. *Education and Treatment of Children, 32*(2), 279-312.
- Cohen, J. D., Servan-Schreiber, D., & McClelland, J. C. (1992). A parallel distributed processing approach to automaticity. *American Journal of Psychology, 2*, 239-269.
- Department of Education and Science (2000). *Learning-support Guidelines*. Retrieved from <http://www.sess.ie/sites/default/files/Learning%20Support%20Guidelines.pdf>

- Fabrizio, M. A., & Moors, A. L. (2003). Evaluating mastery: Measuring instructional outcomes for children with autism. *European Journal of Behavior Analysis*, *4*(1/2), 23-36.
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology*, *104*(1), 206. doi: 10.1037/a0025398
- Gross, T. J., Duhon, G. J., Hansen, B., Rowland, J. E., Schutte, G., & Williams, J. (2013). The Effect of Goal-Line Presentation and Goal Selection on First-Grader Subtraction Fluency. *The Journal of Experimental Education*, *81*, 1-17. doi: 10.1080/00220973.2013.813369
- Haring, N. G., & Eaton, M. D. (1978). Systematic instructional procedures: An instructional hierarchy. In N. G. Haring, T. C. Lovitt, M. D. Eaton, & C. L. Hansen (Eds.), *The fourth R: Research in the classroom* (pp. 23–40). Columbus: Charles E. Merrill Publishing Co.
- Hartnedy, S. L., Mozzoni, M. P., & Fahoum, Y. (2005). The effect of fluency training on math and reading skills in neuropsychiatric diagnosis children: A multiple baseline design. *Behavioral Interventions*, *20*(1), 27-36. doi: 10.1002/bin.167
- Haughton, E. C. (1980). Practicing practices: Learning by activity. *Journal of Precision Teaching*, *1*(3), 3-20.
- Johnson, K. (2008). *Morningside Mathematics Fluency: Math Facts* (Vols. 1-6; curriculum program). Seattle, WA: Morningside Press.
- Johnson, K., & Layng, T. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American psychologist*, *47*(11), 1475. doi: 10.1037/0003-066X.47.11.1475



- Johnson, K., & Layng, T. J. (1994). The Morningside Model of Generative Instruction. In R. Gardner, D. Sainato, J. Cooper, T. Heron, W. Heward, J. Eshleman, & T. Grossi (Eds.), *Behavior analysis in education: Focus on measurably superior instruction* (pp. 173-197). Belmont, CA: Brooks/Cole.
- Johnson, K., & Street, E. M. (2013). *Response to Intervention and Precision Teaching*. New York: The Guilford Press.
- Kubina, R.M., Young, A., & Kilwein, M. (2004). Examining an effect of fluency: Application of letter sound writing and oral word segmentation to spelling words. *Learning Disabilities, 13*, 17-23.
- Kubina, R. M., & Yurich, K. K. L. (2012). *The Precision Teaching Book*. Lemont, PA: Greatness Achieved Publishing Company.
- Lipsey, M. W., & S. M. (2009). Design Sensitivity: Statistical Power for Applied Experimental Research. In L. Bickman, & D. J. Rog (Eds.), *The Sage Handbook of Applied Social Research Methods* (pp. 44-76). SAGE publications, Inc.
- Lin, F. Y., & Kubina Jr, R. M. (2005). A preliminary investigation of the relationship between fluency and application for multiplication. *Journal of Behavioral Education, 14*(2), 73-87. doi: 10.1007/s10864-005-2703-z
- Logan, P., & Skinner, C. H. (1998). Improving students' perceptions of a mathematics assignment by increasing problem completion rates: Is problem completion a reinforcing event? *School Psychology Quarterly, 13*(4), 322. doi: 10.1037/h0088988
- Martens, B. K., & Witt, J. C. (2004). Competence, persistence, and success: The positive psychology of behavioral skill instruction. *Psychology in the Schools, 41*(1), 19-30. doi:10.1002/pits.10135

- McDowell, C., & Keenan, M. (2001). Developing fluency and endurance in a child diagnosed with attention deficit hyperactivity disorder. *Journal of Applied Behavior Analysis, 34*(3), 345-348. doi: 10.1901/jaba.2001.34-345
- Organisation for Economic Co-operation and Development (2010). *Programme for International Student Assessment 2009: What students know and can do – Student performance in reading, mathematics and science.*
- Organisation for Economic Co-operation and Development (2014). *Programme for International Student Assessment 2012: Results in Focus: What 15 year-olds know and what they can do with what they know.*
- Poncy, B. C., Duhon, G. J., Lee, S. B., & Key, A. (2010). Evaluation of techniques to promote generalization with basic math fact skills. *Journal of Behavioral Education, 19*(1), 76-92. doi: 10.1007/s10864-010-9101-x
- Poncy, B. C., Fontenelle IV, S. F., & Skinner, C. H. (2013). Using Detect, Practice, and Repair (DPR) to Differentiate and Individualize Math Fact Instruction in a Class Wide Setting. *Journal of Behavioral Education, 22*(3), 211-228. doi: 10.1007/s10864-013-9171-7
- Shapiro, E. S. (2004). *Academic Skills Problems work-book* (rev.). New York: Guilford Press.
- Skinner, C. H. (2002). An empirical analysis of interspersal research: Evidence, implications, and applications of the discrete task completion hypothesis. *Journal of School Psychology, 40*(4), 347-368. doi: 10.1016/s0022-4405(02)00101-2
- Skinner, C. H., Pappas, D. N., & Davis, K. A. (2005). Enhancing academic engagement: Providing opportunities for responding and influencing students

to choose to respond. *Psychology in the Schools*, 42(4), 389-403. doi:

10.1002/pits.20065

Wechsler, D. (2005). *The Wechsler Individual Achievement Test, Second Edition*.

London: Harcourt Assessment.

Table 1  
*Number of sessions and aims achieved by experimental group*

	Participants in experimental group	Number of sessions completed	Number of fluency aims met <sup>a</sup>
1		16	26
2		25	24
3		16	26
4		28	12
5		23	26
6		22	26
7		23	26
8		28	22
9		21	26
10		23	26
11		24	26
12		24	13
13		25	26
14		30	26

<sup>a</sup>Fluency aims ranged between 50-70 correct responses per minute

Table 2

*Mean performance on each dependent measure at pre- and post-tests for both groups.*

Dependent variable	Time period	Frequency-building	Treatment as usual
		Mean (SD) <sup>a</sup>	Mean (SD) <sup>a</sup>
Mathematical Reasoning <sup>b</sup>	Pre-test	83.2 (10.8)	82.1 (9.57)
	Post-test	88.5(8.08)	82.5(9.6)
Numerical Operations <sup>b</sup>	Pre-test	95.6 (10.1)	92.5 (10.5)
	Post-test	94.4 (7.94)	91.3 (11.3)
Fluency	Pre-test	12.9 (7.65)	11.2 (9.42)
	Post-test	33.1 (13.5)	17.4 (8.85)
Endurance (rate) <sup>c</sup>	Pre-test	12.6 (7.89)	12.8 (10.4)
	Post-test	39.6 (9.59)	20.3 (9.2)
Endurance (On-task) <sup>d</sup>	Pre-test	78 (19)	77.4 (22.1)
	Post-test	99.3 (1.9)	98.6 (2.15)
Stability (rate) <sup>c</sup>	Pre-test	16.7 (9.45)	15.7 (10.4)
	Post-test	45.5 (6.9)	28.8 (7.08)
Stability (On-task) <sup>d</sup>	Pre-test	71.4 (29)	64.7 (27.4)
	Post-test	95.2 (4.78)	92.3 (19.2)
Application	Pre-test	4.93(3.26)	4(3.85)
	Post-test	13.8 (4.86)	11.1 (4.91)

*Note.* All means listed include full sample sizes of 14 for each condition.

<sup>a</sup>Standard deviation. <sup>b</sup>Subtests of the WIAT-II. <sup>c</sup>Rate of correct responding during endurance and stability timings. <sup>d</sup>Observed on-task behaviour during endurance and stability timings.