MASTER THESIS
Learning in Complex Systems
May 2016

Main title:
On the Relationship Between Component and Composite Multiplication Skills and Natural Frequency Growth

Article 1:
Curriculum Based Measurement in the Response to Intervention Framework: A potential tool for effective mathematic interventions

Article 2:
Examining the Relationship Between Component and Composite Frequency and Natural Growth in Basic Multiplication: A systematic replication of Lin and Kubina

Cathrine Berg-Mortensen

Faculty of Health Sciences

Department of Behavioral Sciences

OSLO AND AKERSHUS
UNIVERSITY COLLEGE
OF APPLIED SCIENCES

Acknowledgements

We would first and foremost like to thank our guidance counselor, Børge Strømgren, for his contribution to the study. Your support and guidance in statistical analysis and through the writing process have been highly appreciated. Further, we need to thank the principals and teachers at the participating schools and the students who volunteered to participate. Without them, the current study would have been difficult to conduct. We also need to thank Richard Kubina for permitting us to use the same test-packet, this made it possible to do a systematic replication.

Lastly, we need to thank our families and employers for their patience and understanding during this process.

Abstract

Mathematic proficiency is considered to be a fundamental basic skill, and poor mathematic skills may impact further education, which further could affect national wealth and growth. Results from international tests such as PISA and TIMSS shows that Norwegian students are not doing that well when it comes to do basic skills in arithmetic. Strategies such as *the Knowledge promotion* and *Close to science studies* have been implemented in the Norwegian education system in order to heighten the mathematical competence of the students, but these have not yet seen to be effective. Further are assessment of student proficiency level, and possible learning disabilities traditionally done through norm- and criterion-referenced achievement tests. These tests will rarely show what the potential learning disabilities are, nor will they point to interventions likely to be effective. One system that may facilitate a more preventive approach is Response to Intervention (RtI). RtI aims to be a preventive model through frequent monitoring of student performance, that generates information on potential problems in need of change in the academic environment. Curriculum Based Measurements (CBM) has been found to be highly effective in assessing student performance in reading and writing, and is the preferred measurement procedure in the RtI framework. Precision Teaching (PT) has also been recommended as a supplementary procedure given its frequent monitoring and focus on frequency building, and could as such have the function of both a frequent measurement procedure and as an effective intervention. For an RtI framework to implement effective interventions in mathematics, there is a need for establishing cutpoints or benchmark standards, which could predict further development in the transition from basic to more complex skills and over grade levels. The current study has examined the relationship between component and composite frequency and natural growth in basic multiplication in the grade levels 5–9, and is a systematic replication of Lin and Kubina (2005) with variations.

Keywords: frequency, Curriculum Based Measurement, Precision Teaching, Response to Intervention, component and composite performance, instructional design, skill proficiency, progress monitoring, pinpointing

Curriculum Based Measurement in the Response to Intervention Framework: A potential tool for effective mathematic interventions

Cathrine Berg-Mortensen

Oslo and Akershus University College of Applied Sciences

Abstract

Mathematic proficiency is considered to be a fundamental basic skill, and poor mathematic skills may impact further education, which further could affect national wealth and growth. Thus, Mathematics is considered to be a core subject in the Norwegian compulsory education. Norway spent, in 2014, 65 billons in middle and high school, this equals to 13% of the national budget, together with strategies such as, increased number of hours devoted to core subjects and increased emphasis on basic skills, to try to increase students' academic performance. However, despite these strategies the trend in mathematic proficiency is in decline. In order to achieve a precise knowledge about each student's learning, schools could profit from setting benchmarks of what progress looks like each year. Hence there is a need for more specific and effective screening procedures, which are aimed at being predicable in regards to students' future academic outcome. Curriculum Based Measurements (CBM) have been found to be highly effective in assessing student performance in reading and writing, and are the preferred measurement procedure in the Response to Intervention (RtI) framework. RtI aims to be a preventive model through frequent monitoring of student performance, which generates information on potential problems in need of change in the academic environment. For an RtI framework to implement effective interventions in mathematics, there is a need for establishing cutpoint or benchmark standards, which could predict further development in the transition from basic to more complex skills and over grade levels.

Keywords: Curriculum Based Measurements, Response to Intervention, cutpoints, Precision Teaching

Mathematic proficiency is considered to be a fundamental basic skill (Kunnskapsdepartementet, 2013; Lin & Kubina, 2005). There is good reason for this as poor mathematic skills may impact further education, which again could affect national wealth and growth (Hanushek & Kimko, 2000). Also, there is a strong societal consensus that all students should attain basic skills in mathematics, as it is important for the individuals' ability to attain a successful life (Murnane, Willett & Levy, 1995). This is not to say that all students can or must become scientists, economists or engineers, but all should and deserves to be able to perform skills that are necessities of living and participating in society. Thus mathematics, along with reading, writing, and natural sciences are considered core subjects in the Norwegian compulsory education system. In 2014, 13% of the national budget was spent on education, 65 billion in middle and high schools, and a total of 2588 hours was devoted to the core subjects (Haugberg, 2014).

Since *The Knowledge Promotion* was implemented in august 2006, mathematics has been a priority (Kunnskapsdepartementet, 2013). *The Knowledge Promotion* aimed at reforming the education system in order to increase the academic achievements amongst students. Aside from spending money on education, this aim was to be reached through different strategies such as: (1) giving schools more autonomy in selecting teaching methods and instructional design, (2) increased number of hours devoted to core subjects in middle and high school, and (3) increased emphasis on basic skills. However, despite these strategies the trend in mathematic proficiency is in decline (Haugberg, 2014). Spending more money and devoting more time to the core subjects, than other OECD (*Organization for Economic Cooperation and Development*) countries does not seem to pay off. In fact, the amount of students in need for special education increases with students' age and grade level, from 3.6% of students in the first grade level, to 11.1% of students in the 10th grade level (Utdanningsdirektoratet, 2015a).

The 2011 results of the international assessment, *Trends in mathematics and sciences studies* (TIMSS) illustrated that the biggest challenge Norway faces in education, primarily mathematics, is poor student performance in arithmetic, mainly algebra and numeracy (Grønmo et al. 2012). Several studies, including TIMSS and *Program for International Student Assessment* (PISA) indicates that low student performance in mathematics is an international problem (Kjærnsli & Olsen, 2013). Lin and Kubina (2005) report that US students perform similar to students in countries like Romania, Spain and Iran, who do not spend billions of dollars on education. They still perform at the same level as US students in mathematics.

Similar trends are seen in the Norwegian PISA and TIMSS tests: 4th graders are improving, while 8th graders show a decline in arithmetic skills. Also, other OECD countries spend less money and have fewer hours of the core subjects and still score at the same level as Norwegian students (Haugberg, 2014). The Norwegian situation regarding student performance has not improved since 2003. In 2015, 14.5% of Norwegian 10th graders failed the final exam in mathematics, a 5.6% increase from the school year of 2010/2011 (Utdanningsdirektoratet, 2015b). *The Knowledge Promotion* strategies (spend money, more school autonomy regarding teaching methods and instructional design, and more time spent on core subjects including mathematics) does not seem to be effective towards its goals.

Consequently, there has been a debate regarding how student proficiency in mathematics should be improved. In response, *The Ministry of Education and Research* (Kunnskapsdepartementet; KD) are planning to implement a five-year national strategy called *Close to Science studies* (Kunnskapsdepartementet, 2015). The overall aim is, again, to reduce the number of students below benchmark in mathematic proficiency, and ensure that the majority of students can perform at a higher and more advanced level in natural sciences. The aim is to be achieved through (1) renewal of the teaching methods and instructional design,

(2) introducing natural sciences in kindergarten, and (3) demanding an overall higher proportion of teachers with extra or specialized training in mathematics and science.

Both the KD strategy for *Close to Science studies* and the previous debate concerning mathematics proficiency improvement discussed some interventions proven to have a low desired effect (Hattie, 2009). These include (1) increased number of teachers/adults in schools, (2) smaller class sizes, (3) higher admission standards for teaching academies, (4) increased training/specialization of teachers, and (5) an increase from 4 to 5 years of teacher education.

They are all ranged as having a low educational effect size; Cohen's *d* = .40 or less. That is, if they are not combined with other empirical supported interventions (Hattie, 2009). Strategy 1 and 2 spells "spend more money", which is probably not going to improve proficiency rates (Hattie, 2009; Lin & Kubina, 2005). Neither are strategies 3, 4, and 5. Hattie (2009) contends that the main problem is not that teachers lack appropriate teaching skills, but that there is a large variance in their effect on students' learning, meaning that teachers do not know how their students profit from their teaching. Actually, these strategies are assessed by Hattie (2009):

> Any inspection of the policies of state or federal government, however, would show that there are few policies that directly affect teaching. Most policies are about structural issues such as resources, smaller class sizes, choice (of whom you want to send your children to school with), curriculum, and tests and high stake assessment. It is rare to find a policy that relates to learning. (p. 244)

This is not to say that structural issues are not important, according to Hattie (2009) they may increase the probability that active teaching strategies will be successful. Hattie (2009) lists a number of active teaching strategies that have proven effective. However, it is not enough to focus on a specific strategy or a script in order to teach successfully:

> It is not a particular method, nor a particular script, that makes the difference; it is attending to personalizing the learning, getting greater precision about how students are progressing in this learning, and ensuring professional learning of the teachers about how and when to provide different or more effective strategies for teaching and learning. (Hattie, 2009, p. 245)

Hattie (2009) suggest that in order to achieve a precise knowledge about each student's learning, schools need to set benchmarks of what progress looks like each year, in addition to an awareness of what individual students are knowing and thinking. Furthermore, in order to secure active and efficient teaching strategies schools must collect effect sizes toward these benchmarks in order to determine what works best for whom, when and why. One way to achieve this is through assessments and benchmark tests for all students.

Some Commonly Used Assessments and Benchmark Tests

Norm-referenced achievement tests are commonly used in the education system. The analysis unit of these tests are how well students perform, compared to other students, measured toward an expected level of performance based on age, grade or gender (Shapiro, 2011). PISA and TIMSS are examples of typical norm-referenced achievement tests, which are commonly used as comparison of general student performance at an international level, as well as an illustration of national student proficiency in specific grade levels. Another widespread method for assessing student proficiency is through criterion-referenced achievement tests, which are used to identify individual student proficiency within a targeted skill area. Measurements of criterion-referenced achievement tests are based on individual student performance to a predetermined absolute standard (Shapiro, 2011).

The Norwegian Directorate for Education and Training (Udir) has recently developed *Supportive learning tests* in mathematics for grade levels 5–7 (primary) and 8–10 (secondary school). They are designed to measure student proficiency in basic math skills. However, each school can leave it to their own discretion to use these tests or not, and should they do so, they are free to decide when during the academic year the tests are to be administered (Kunnskapsdepartementet, 2015). Udir has also implemented *National tests* to examined student proficiency in grade level 5, 8 and 9. These tests are aimed to serve as a diagnostic tool for monitoring progress and differences across schools. *National tests* are administered at the start of each grade level (5^{th}, 8^{th} and 9^{th}) and results yield how well students in one school scores in comparison to another school. Further, they are to be used locally for teachers as an indication of individual student proficiency in core subjects (Utdanningsdirektoratet, 2016). *National tests* and *Supportive learning tests* are varieties of criterion-reference tests, in addition to yield overall student performance, i.e. norm-referenced achievement test.

Both norm-referenced and criterion-referenced achievement tests have some advantages. Criterion-referenced achievement tests can be used to identify struggling students, which can generate earlier intervention in order to increase the student's learning. Norm-referenced achievement tests yield information about larger groups of students and overall performance, which can indicate current and past performance developments (Shapiro, 2011).

Despite those advantages, norm- and criterion-referenced achievement tests alone may not be an adequate foundation for all academic decisions. In assessing student proficiency towards expected outcomes they have a function. However, in terms of assessing the effectiveness of the current curricula and teaching strategies on student learning, and monitoring individual student performance over time to identify problem areas, they fall short (Shapiro, 2011). Identifying potential learning disabilities and making decisions regarding possible effective interventions, based on results attained through procedures that measures the discrepancy between the expected proficiency level and the attained score by the student, raises many questions. Mainly concerning the procedures' validity and reliability, and they have been found to lack empirical support (Shapiro, 2011; VanDerHeyden & Burns 2010; Barnett et al., 2007).

So, despite the existence, and to some degree the use of, assessment procedures suggested by Hattie (2009), teachers may still not know whether or how their students profit from their teaching. Thus, there is a need for reliable and valid measurement procedures in order to accurately assess student performance and to generate more specific interventions.

This calls for a schoolwide system of regular assessment of student proficiency and performance throughout the schoolyear, and through all grade levels. More accurate and reliable screening procedures, termed brief universal screening procedures, occurred as a response to the traditional procedures based on the discrepancy model. The overall goal of these procedures are to assess both student proficiency and the effectiveness of the current curricula and/or teaching strategies upon student learning. The common perception of more direct measurements procedures is that one "…should test what one teach" (Shapiro, 2011, p. 27). Furthermore, based on the assessment of student progress, necessary efficient interventions need to be put in place in a timely fashion. A schoolwide systems such as *Response to Intervention* (RtI), may be helpful in this effort.

Response to Intervention

RtI originated from multiple areas of research in the field of applied behavior analysis, such as; Precision teaching (PT), Curriculum based Measurements (CBM) and diagnostic prescriptive teaching, amongst some (VanDerHeyden & Burns, 2010; Johnson & Street, 2013; Shapiro, 2011). RtI is defined as a framework for decision making, consisting of integrated methods and procedures aimed at effectively identifying academic areas in need of improvement (VanDerHeyden, Witt & Gilbertson, 2007; Johnson & Street, 2013). Effects of implementing RtI is found to decrease the rate of students developing learning disabilities and increasing the probability of positive academic outcomes for all students (Deno et al., 2001;

Shapiro, 2011; Johnson & Street, 2013). This is done by emphasizing on conducting universal screening procedures frequently at all stages in the education system, and targeting students at risk for academic deterioration, and such in need for more intensive interventions (VanDerHeyden & Burns, 2010).

The overall goal for RtI is to increase the probability of positive academic outcomes for all students, including struggling learners. As a framework, RtI aims at being a preventive model and is often, and best, described as a service delivery system consisting of different tiers. How RtI operates at the different tiers is illustrated through a triangular model with a wide bottom and narrow top. Tier 1 refers to the bottom level, and constitutes students who benefits from an empirically supported class-wide curriculum. Tier 1 students are generally assessed based on general outcome measurements, for instance a brief universal screening procedure such as CBM (VanDerHeyden & Burns, 2010; Johnson & Street, 2013; Shapiro 2011). These brief screening procedures are administered approximately three times a year (fall, winter and spring) and yield information on the overall student performance, and also the effectiveness of the current class-wide curriculum (VanDerHeyden & Burns, 2010). The unit for analysis in brief screening procedures, such as CBM, is the response frequency which is calculated by counting correct and incorrect responses per minute. The duration of the measurement interval varies, but in CBM it is usually from 1-3 minutes in reading and writing, and 2-8 minutes in mathematics (Shapiro, 2011).

The developers of CBM aimed at establishing a measurement procedure that could be; (1) an effective tool for teachers, (2) accurate in regards to student proficiency and growth, (3) account for the effect of current curricula and teaching strategy and (4) yield a more accurate and valid foundation for the implementation of academic decisions (Deno et al., 2001). As educational decisions in a RtI framework need to be data-based, in order to identify potential changes in the academic environment that could accelerate student learning. CBM proposes as an efficient and accurate procedure to account for data-based decisions within the framework (Shapiro, 2011). CBM as a brief universal screening procedure aims at assessing basic skills in spelling (CBM-S) reading (CBM-R), writing (CBM-W) and mathematics (CBM-M) (Christ et al., 2008; Shapiro 2011). The predictive value of CBM in regards to future student performance has been well documented in the domain of reading and writing (Shapiro, 2011). Current research in CBM-M shows potential for both the assessments of student growth in basic mathematical skills and the current curricula effectiveness upon student learning (Christ et al., 2008). By conducting CBM three times a year, teachers and instructors can more accurately assess if the given curricula are effective, and identify students who are in need for more intense monitoring and may not benefit from the current curricula, and thus are reassign to Tier 2 or 3 depending on their needs.

Tier 2, refers to groups of students with similar academic problem areas and the interventions are group-based. Students in Tier 2 are detected by the Tier 1 brief universal screening procedure. Data from these screenings are aimed to predict future proficiency levels in a given core subject. Given poor student performance, interventions are put in place to enhance the probability of students catching up and as such, could continue to follow the core-curricula. If the interventions not produce the expected result, further assessments may be needed in order to detect which skills are lacking or where the curricula fail to accelerate student learning. By pinpointing were the problem is located, more efficient interventions can be put in place in a timely fashion that can remedy the skill deficiency (Some of these effective interventions will be given a brief description later in the current article) (Codding et al., 2009; Codding, Burns & Lukito, 2011).

Should interventions at Tier 2 not prove to be effective for the students in question, more fine grained assessment procedures may be needed in order to pinpoint interventions that will accelerate academic learning. These procedures may be similar to those necessary for students in Tier 3, but also principles derived from PT may be necessary. PT is based on basic behavioral principles and was developed by Ogden Lindsley in the 1960's as an alternative approach to the current academic interventions for students with learning disabilities. PT is a measurements system that incorporates progress monitoring, pinpointing and intervention analysis and hence benefits a RtI framework. Frequency building and its implications are the main objective in a PT procedure. Five specific steps comprise the process of PT, (1) pinpointing the learning objective, (2) develop materials and procedures (3) time and count response frequencies (4) charting and (5) data-based decisions (Johnson & Street, 2013; Kubina & Yurich, 2012). Students who do not benefit from class-wide curriculum and group administered interventions are found in Tier 3, were interventions are administered at an individual level. Data in Tier 3 is collected through frequent progress monitoring and gives session-by- session, or at least daily information about both progress and intervention effects (VanDerHeyden & Burns, 2010; Johnson & Street, 2013).

The diagnostic value of RtI lies in its key components (1) brief universal screening, (2) empirical supported core curriculum, (3) progress monitoring, (4) increasingly intensive interventions and (5) data-based decisions (VanDerHeyden & Burns, 2010; Johnson & Street. 2013). Brief universal screening procedures are administered to all, or groups of students, to identify students at risk for academic failure and also to assess the effectiveness of the core-curriculum. Students at risk for academic failure are more frequently monitored (progress monitoring) and receives more intensive interventions. Data from brief universal screening and progress monitoring enables teachers and instructors to more accurately assess the academic environment and establish effective intervention were it is needed (Olson et al, 2007).

RtI is in combination with procedures such as CBM and PT, influenced by Diagnostic prescriptive teaching (VanDerHeyden & Burns, 2010). Diagnostic prescriptive teaching evolved as an alternative procedure for addressing learning disabilities, and aimed to make data-based decisions for implementing educational interventions. The procedure of Diagnostic prescriptive teaching involved assessing students' performance and prescribing interventions in regards to the outcome of the assessments. As a procedure the diagnostic prescriptive teaching approach did not succeed, but it has influenced the development of more preventive and diagnostic instructional intervention programs and measurement procedures, such as RtI (VanDerHeyden & Burns, 2010).

In the next section Brief Universal Screening procedures, such as CBM, and effective interventions at the different tiers in the RtI framework will be highlighted, subsequently will there be a brief description on how PT supplements RtI. Further will CBM as a potential tool for implementing effective mathematical interventions in a RtI framework be discussed.

RtI at Tier 1

Core Curriculum as an Effective Intervention

The RtI framework holds that the core-curriculum needs to be comprehensive, i.e. account for all parts of the given subject, comprise of effective and empirically supported methods and should be monitored at a frequent basis (three times a year). Further should a core curriculum in a given subject be correlated with evidence that shows its effectiveness in reaching age/grade/district or area goals, in sum secure that students master a given skill before further progression (VanDerHeyden & Burns, 2010). The core-curriculum in Tier 1 is the intervention in effect and brief universal screening procedures are aimed at measuring the effectiveness of the given curriculum and student proficiency level (Johnson & Street, 2013).

Core-Curriculum administered in regular schools are usually not monitored at a regular basis or measured towards its effect on student learning (Johnson & Street, 2013; Shapiro, 2011), hence the frequency of international and national assessments (Grønmo et al., 2012; Kjærnsli & Olsen, 2013; Utdanningsdirektoratet, 2016; Kunnskapsdepartementet, 2015). Infrequent testing could depict results that not necessary reflect the given academic environment, which could generate ineffective core-curricula and teaching strategies. Which could lead to an increase of students at risk for academic deceleration, and further make it more problematic to assess the specific academic areas in need of improvement.

As the core-curriculum in a RtI framework need to meet specific criteria (consist of empirical supported methods and be comprehensive) to be set in effect, and is prone to frequent monitoring of its effect on student learning. It might be more likely to succeed in enhancing student proficiency, and further facilitate positive academic outcome for the general student (Johnson & Street, 2013).

Brief Universal Screening

Brief universal screening procedures and progress monitoring are, as previously mentioned, considered to be more accurate and precise measurements procedures than norm- and criterion-referenced achievement tests, and establish the foundation for data-based decision in a RtI framework (Barnett et al., 2007; Olson et al., 2007; VanDerHeyden & Burns, 2010; Johnson & Street, 2013). CBM is the preferred applied screening procedure in RtI (Shapiro, 2011; Johnson and Street, 2013; VanDerHeyden & Burns, 2010), both as an universal screening procedure and as progress monitoring, given that it serves multiple areas of identification (identifies struggling learners, the effectiveness of the curricula and provides a baseline for establishing grade and individual academic goals). At grade level (Tier 1) CBM is often administered three times a year, normally during the fall, winter and spring to assess student growth and to obtain normative data for a specific school (Shapiro, 2011). Universal screening procedures are always conducted before an implementation of a RtI framework, as it will identify where the potential problem (student skill deficit or ineffective curriculum) is located (Johnson & Street, 2013).

Results from CBM enables teachers and instructors to reassign students to other tiers given their outcome. These data can also provide a baseline for establishing more accurate grade level goals through the progression of the curriculum (Johnson & Street, 2013). Further it can yield results indicating the effectiveness of the current curricula, given that a large proportion of the students failed to reach the grade goal might correspond to the curricula or teaching strategies in effect, and not student proficiency (VanDerHeyden & Burns, 2010). If students need reassignment to a different tier, the interventions methods and screening frequency will intensify.

RtI at Tier 2

Effective Interventions

Interventions aimed at students in Tier 2 are, in similarity to the core-curriculum, required to have empirical support for its effectiveness upon student learning and be possible to be carried out in group-formats. Empirical supported interventions found to have an effect on student progression and enhance learning in basic skills, are drill and practice methods, and methods were students monitor their own progress, set their own goal and apply interventions (Codding et al., 2011). Copy-cover-compare (CCC), Self- Monitoring and Peer Tutoring (Codding et al., 2009) are examples of interventions with empirical support for its effect upon student learning. The CCC method has strong scientific support and is applicable in many different settings and over multiple skills, but has shown to be especially effective in the areas of math and spelling. The process of CCC usually consists of three steps. Where the first step involves that a stimulus is presented and the student is asked to copy it, e.g. a student may be presented with a math fact and the correct answer to that mat fact, and asked to copy the correct answer. Step two involves covering the stimulus (e.g. the correct answer) and then copying it by memory. The last step consists of comparing the original stimulus to the copied stimulus (Joseph et al., 2011).

Self-Monitoring has also proven to be an effective tool for interventions aimed at changing academic behavior, and is founded on the basic principle of measuring and evaluating a target behavior to a predetermined aim. The process composes of two main components, record and measure. Where the students record and measure their own behavior and evaluate the recorded behavior to a predetermined goal. There are different variations of the Self-Monitoring procedure, but it usually involves steps as; defining target behavior, deciding a monitor schedule, continuously evaluation and fading of the procedure (Wright, 2013). Another intervention proven to be effective is Peer-Tutoring. This method comprises of pairing high performing students with lower performing students, in a tutor-tutee relationship. The pairing can be arranged in groups (class wide) or by pairing two students by age or proficiency level. The procedure has proven to be effective in increasing both self-confidence and self-efficacy as well as student proficiency (Hott, Walker & Sahni, 2012).

Interventions, like these, are applied to supplement the core curriculum and to prevent students from continuous decelerating as the core-curriculum advances. As intervention increases so does the monitoring of student performance and screening procedures (VanDerHeyden & Burns, 2010; Shapiro, 2011; Johnson & Street, 2013).

Progress Monitoring

Progress monitoring in Tier 2 also requires empirical support for its sensitivity to identify the academic environment (student performance and curriculum effect). Further it need to be brief and easy to administer (Shapiro, 2011). CBM has been identified as an effective measurement procedure at both Tier 1 and Tier 2 given its sensibility to the current curricula, interventions and student performance (VanDerHeyden & Burns, 2010). The recommended frequency of progress monitoring at Tier 2 ranges from every other week (Shapiro, 2011) to a couple times per month (Johnson & Street, 2013). There is usually not a need for more than 4–6 screenings before a decision regarding student performance or the effectiveness of interventions can be made (Shapiro, 2011). Data from screening procedures in Tier 2 enables teachers and instructors to evaluate the effectiveness of the current intervention and reassigning students to another tier. Given that students show a lack of improvement in Tier 2, interventions and progress monitoring increases and the students are referred to Tier 3.

RtI at Tier 3

Effective Interventions

Effective interventions in Tier 3 are required to have empirical support and be highly efficient in targeting basic skills and problem areas in the students' academic behavior repertoire. The interventions are usually administered in a 1–1 format, and focuses on specific skills. One systematic measurement and intervention system found to be highly efficient, at both group (Tier 2) and individual (Tier 3) level, to enhance student learning is PT (Johnson & Street, 2013). The overall goal of PT is frequency building of academic skills to mastery. Mastery in PT refers to the concept of fluency, and fluency is defined as accuracy plus speed (Binder, 1996; Johnson & Street, 2013; Lin & Kubina 2005; Strømgren, Berg-Mortensen &Tangen, 2014). The academic outcome of performing a skill to fluency refers to that the skill can be maintained in the future (maintenance), over longer period of times (endurance), continued in distractive conditions (stability), applied in other situations (application) and emerge in to more complex behavior repertoires (generativity) (Johnson & Street, 2013).

As such, PT can, by emphasizing frequency building of component skills, enhance student proficiency in skills necessary for further progression through a curriculum and be an effective measurement procedure for students in Tier 2 and 3. Further is the notion of frequency building of basic skills to mastery in PT, a potential effective strategy to be applied for Tier 3 students (Johnson & Street, 2013). As there is a strong societal and scientific consensus that basic component skills should be mastered before moving on to more complex composite skills (Wong & Evans, 2007; Johnson & Street, 2013; Kunnskapsdepartementet, 2013; Binder, 1996; Strømgren et al., 2014; VanDerHeyden & Burns, 2009).

Progress Monitoring

By assessing student proficiency through brief universal screening procedures and progress monitoring, both CBM and PT can be effective measurement procedures, as they are aimed to identify changes that could enhance student learning in Tier 2 and 3 by providing continuous progress monitoring. As CBM is a measurement procedure found to produce valid and reliable outcome prediction in some core subjects (reading and writing) (VanDerHeyden & Burns, 2010), PT can function as a supplementary screening procedure as it incorporates daily charting of progress (Johnson & Street, 2013). As a result of daily charting (see for instance Johnson and Street, 2013 or Kubina and Yurich, 2012 for more information regarding the charting procedure), the interventions can alter at a more frequent level to more accurately address the current proficiency level of the individual student.

Given the combination of drill and practice procedures and progress monitoring, PT can be an effective tool for assessing students' performance and pinpointing where effective interventions might be needed (Johnson & Street, 2013). Procedures from PT, both in regard to progress monitoring and frequency building, can be a good match for the RtI framework (Johnson & Street, 2013).

PT as a Supplementary Procedure in a RtI Framework

Johnson and Street (2013) emphasize that PT is highly compatible with RtI as it can supplement screening procedures in the different tiers and be a precise placement tool in the process of reassigning students for Tiers 2 and 3 by providing more fine-grained assessments of skill proficiency and efficient interventions. PT also put emphasis on pinpointing and frequent charting as crucial steps which could make teachers and instructors able to more accurately pinpoint potential factors that could increase student proficiency. In addition, does the daily charting offer frequent data that indicate student growth by generating a day-to-day analysis of the current effect of the intervention and as such the intervention can alter in correlation to the daily analysis. Another beneficial component PT displays, is rapid frequency building. Given that the students show a lack of basic and component skills, PT could supplement other intervention to increase the probability of students reaching benchmark, and as such can follow the core-curriculum. In this context PT can supplement progress monitoring in Tier 2 and 3, and further be an effective intervention to apply in Tier 3 (Johnson & Street, 2013)

CBM as an Effective Assessment Tool for Mathematical Interventions in the RtI Framework

As the norm- and criterion- referenced achievement test can be regarded as somewhat limited and faces certain challenges, i.e., only addresses student proficiency towards a standardized level. CBM can be a more precise tool for assessing student proficiency, and the effectiveness of the current teaching strategies in mathematics. CBM has showed to be effective in both assessing student proficiency and the applied curricula in the core subjects of reading and writing (Deno et al., 2001; VanDerHeyden & Burns, 2010; Shapiro, 2011). By founding assessments on the local curricula, the CBM procedure holds to be more susceptible in detecting student proficiency in core subjects and to more accurately identify struggling learners. However, there is a current limitation that needs to be addressed before a unified consensus regarding CBM's potential effectiveness in a RtI framework for mathematics is made. Given that CBM aims at assessing student proficiency within a curriculum, there is a need for more accurate knowledge on cutpoints, i.e. benchmarks for academic performance developments in mathematics, that could predict future student outcome, as a curricula progresses from basic to more complex mathematical skills. Given that CBM-R and CBM-W have found crucial cutpoints in reading and have good empirical support for its predictive value in regards to student performance within their respective core-subjects. The same might be achieved in the domain of mathematics. Brief experimental analysis (BEA), CBM and PT amongst some have tried to identify scores or frequencies, that could predict further growth and possibly the emergence of more complex behavior repertoires (e.g., generativity) in the transition from basic to more advanced mathematical skills (VanDerHeyden & Burns, 2010; Johnson & Street, 2013; Haughton, 1972).

VanDerHeyden and Burns (2009) applied BEA assessments to examine its potential for developing reliable cutpoints or benchmark scores in mathematics, which could predict further deceleration or growth as curricula advances. Results from their study showed that students who reached the frequency aim in basic skills were more likely to perform at the frequency aim in more complex skills. Research within the field of PT have investigated if a given frequency in component skills can predict further growth in more complex skills (Haughton, 1972; Binder, 1996; Johnson & Street, 2013). Different frequency aims have been discussed throughout the PT literature. Haughton (1972) referred to writing single digits at a frequency of 80 or more per minute, and computing 40–50 basic math fact per minute, as potential frequency aims that could predict further growth as the curricula progresses. A lower frequency would predict deceleration. Other studies have found approximately the same indication, that a frequency of 30 or fewer basic math facts per minute gives an indication of deceleration as the curricula advances (Johnson & Street, 2013; Wong & Evans, 2007; Strømgren et al., 2014). Assessments of mathematical proficiency through CBM-M have found CBM-M to be a reliable measurement of fluency in group-based assessments (VanDerHeyden & Burns, 2008). Other studies within brief universal screening procedures and PT have showed that attained mathematical scores in lower grades corresponds with mathematical scores in higher grades (Deno et al., 2001) and that proficiency in basic skills is correlated with proficiency in more complex skills (Lin & Kubina, 2005; Strømgren et al.

2014). As research shows, there is a need for further studies, that aims to investigate the validity and reliability of CBM-M, in its ability to establish cutpoints in the domain of mathematics before a unified conclusion about its predictive value can be drawn (Christ et al., 2008).

However, previous research has showed that when student performance is measured repeatedly and systematically it yields a more realistic picture of overall student academic performance, and enhance the probability of detecting struggling learners at an earlier time, which in turn increases the student's probability of not falling further behind in the curriculum (Shapiro, 2011). Which strengthens the notion that brief universal screening procedures, such as CBM can produce reliable and valid cutpoints or benchmark standards, that could predict further student performance through a curriculum. In order to achieve more knowledge on potential cutpoints in mathematics through CBM, there is a need for more studies on students' natural performance growth over age and grade level. Measuring student proficiency in mathematics at a frequent level, and based on the current curricula in effect over time, could yield information on both potential student progress as the curriculum advances, and the effectiveness of the curricula and/or teaching strategies in effect.

Given more accurate knowledge on the cutpoints associated with student growth, it could be an effective and precise measurement procedure to apply in a RtI framework for mathematics. As it could identify students at risk for academic failure and arrange more effective interventions to be applied where it is needed.

Summary

International tests such as PISA and TIMSS illustrates that student performance in mathematics not corresponds with strategies (e.g., amount of money spent and the amount of teachers in schools) arranged by the Norwegian government (Kjærnsli & Olsen, 2013; Grønmo et al., 2012). In a national perspective PISA and TIMSS depict that although 4th

graders seem to improve in mathematical proficiency, 8th graders are in decline. These results indicate that something happens in the transition from basic mathematical skills learned in the 4th grade, to more complex mathematical skill acquisition expected in the 8th grade. Which corresponds with previous research conducted to investigate growth from basic to complex skills (Lin & Kubina, 2005; VanDerHeyden & Burns, 2009). As they have demonstrated that student performance in lower grades corresponds with student performance as the curricula progresses through grade levels. Results from these tests are further measured toward a predetermined standard based on grade level, and give little or no information regarding the potential targeted problem areas students face, in the progression from basic to more complex mathematical skills. To assess the current discrepancy between students showing improvement in the 4th grade and students in decline in the 8th grade. A possible solution could be more frequent measurements through brief universal screening, such as CBM. Administering CBM at a frequent level could be a more susceptible approach in assessing when students shows signs of struggling, and further enables more effective interventions, which could generate a more preventive approach (VanDerHeyden & Burns, 2010; Johnson & Street, 2013; Shapiro, 2011).

Another potential problem with international and national test results is that they are used as a basis for forming educational decisions and thus given the limitations with such test, the strategies developed may not have the intended effect upon student learning. As mentioned in the introduction The Norwegian government aims at introducing intervention (Kunnskapsdepartementet, 2013/2015; Utdanningsdirektoratet, 2016) that have shown to have little effect if not combined with other empirically supported interventions (Hattie, 2009). Given that decisions for potential effective interventions are based on normative data, they may not correlate with the actual conditions in need of change and as such, they may not have the expected impact upon the academic environment (Johnson & Street, 2013). More accurate and frequent screening procedures based on assessing student performance within a given curriculum could be a more comprehensive approach. By collecting more reliable and valid data could enhance the probability of implementing effective interventions (Johnson & Street, 2013; VanDerHeyden & Burns, 2010; Shapiro, 2011).

A framework found to incorporate elements, such as brief universal screening procedures and effective interventions is RtI. Given its preventive approach for all core subject in which; (1) brief universal screening procedures, such as CBM, that has proven effective in assessing both student proficiency and the current curricula administered in a subject (Deno et al., 2001; VanDerHeyden & Burns 2009/2010; Johnson & Street, 2013; Shapiro, 2011), (2) Progress monitoring through CBM and PT, (3) effective interventions such as, CCC, self-monitoring, peer tutoring and PT. Which have shown to be effective in targeting problem areas and increasing student proficiency in basic skills (Joseph et al, 2011; Codding et al., 2009; Codding et al., 2011; Johnson & Street, 2013; Binder, 1996), (4) data-based decisions derived from brief universal screenings and progress monitoring and (5) an empirically supported core curriculum, play vital parts (VanDerHeyden & Burns, 2010). Given these elements RtI proposes to be effective in enhancing student performance (Joseph et al., 2011; Wright, 2013; Hott et al., 2012; Johnson & Street, 2013; Shapiro, 2011) and reduce the amount of students developing learning disabilities (VanDerHeyden & Burns, 2010). Given those implications, the RtI framework could be an effective approach in enhancing student performance in mathematics.

An important objective that need further research before implementation of RtI in mathematics can be done successfully, have been described as the need for establishing cutpoint or benchmarks that could predict further development through a progressive curriculum. Different studies in PT, CBM and BEA have tried to identify frequency aims that holds a predictive value for further academic advancement. At the moment, most studies have concluded that there is a correlation between levels of frequency in basic skills and attained levels of frequency in complex skills (VanDerHeyden & Burns, 2009/2010; Wong & Evans, 2007; Lin & Kubina, 2005). A possible solution for identifying cutpoints in mathematics is assessing natural frequency in mathematics by conducting CBM in lower grades and following those grades as they progress through the education system.

Given an increased knowledge of students' natural frequency in mathematics through an education system, may help establishing cutpoints that could predict future student academic outcome with and without effective interventions. This information could generate more accurate identification of struggling students, and further increase the possibility of intervening at an earlier time, and hence possibly turn the trend of poor student proficiency in mathematics.

References

Barnett, D. W., Hawkins, R., Prasse, D., Graden, D., Nantais, M. & Pan, W. (2007). Decision-Making validity in reponse to intervention. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention. The science and practice of assessment and intervention*. (pp. 106–116). New York NY: Springer Science+Business Media.

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163–197. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2733609/pdf/behavan00020-0021.pdf

Christ, T. J., Scullin, S., Tolbize, A. and Jiban, C. L. (2008). Implications of recent research. Curriculum-Based measurement of math computation. *Assessment for Effective Interventions, 33*(4)198–205. doi: 10.1177/1534508407313480

Codding, R. S., Hilt-Panahon, A., Panahon, C. J. & Benson, J. L. (2009). Adressing mathematics computation problems: A review of simple and moderate intensity interventions. *Education and Treatment of Children, 32*(2)279–312. doi: 10.1353/etc.0.0053

Codding, R. S., Burns, M. K. & Lukito, G. (2011). Meta-Analysis of Mathematic Basic-Fact Fluency Interventions: A Component Analysis. *Learning Disabilities Research & Practice, 26*(1)36–47. doi: 10.1111/j.1540-5826.2010.00323.x

Deno, S. L., Fuchs, L. S., Marston, D. & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4)507–524. Retrieved from http://web.b.ebscohost.com.ezproxy.hioa.no/ehost/detail/detail?sid=4f18348b-8215-4b52-bb22-

a815bc936b24%40sessionmgr103&vid=0&hid=107&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ%3d%3d#AN=5962052&db=aph

Grønmo, L. S., Onstad, T., Nilsen, T., Hole, A., Aslaksen, H. & Borge, I. C. (2012). *Framgang, men langt fram. Norske elevers prestasjoner i matematikk og naturfag i TIMSS 2011 [Progress, but still a long way to go. Norwegian students' 2011 TIMSS performance in mathematics and science]*. Oslo: Akademika.

Hanushek, E. A. & Kimko, D. D. (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review, 90*(5)1184–1208. Retrieved from http://www2.econ.iastate.edu/classes/econ520/Huffman/documents/SchoolingLaborForceQualityandtheGrowthofNations.pdf

Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achivement*. London: Routledge.

Haugberg, T. (2014). *The Education Mirror. Facts and analysis of kindergarten, primary and secondary education in Norway*. Retreived from http://www.udir.no/globalassets/upload/rapporter/educationmirror/the-educationmirror_english.pdf

Haughton, E. (1972). Aims-growing and sharing. In J. B. Jordan & L. S. Robbins (Eds.), *Let's try doing something else kind of thing: Behavioral principles and the exceptional child. A report from the invisible college conference on application of behavioral principles in exceptional child education, March, 1971*. Arlington, VA: The Coucil for Exceptional Children. Retrieved from http://www.fluency.org/Haughton_Aims_1972.pdf

Hott, B., Walker, J. & Sahni, J. (2012). *Council for Learning Disabilities*. Retrieved from http://www.council-for-learning-disabilities.org/peer-tutoring-flexible-peer-mediated-strategy-that-involves-students-serving-as-academic-tutors

Johnson, K. & Street, E. M. (2013). *Response to Intervention and Precision Teaching. Creating Synergi in the Classroom*. New York: The Guilford Press.

Joseph, L. M., Konrad, M., Cates, G., Vajcner, T., Eveleigh, E. & Fishley, K. M. (2011). A meta-analytic review of the cover-copy-compare and variations of this self-management procedure. *Psychology in the Schools, 49*(2)122–136. doi: 10.1002/pits.20622

Kjærnsli, M. & Olsen, R. V. (2013). *Fortsatt en vei å gå. Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012 [Still a way to go. Norwegian Students' proficiency in Reading, Mathematics and Science Literacy in the PISA Assessment 2012]*. Oslo: Universitetsforlaget.

Kubina, R. M. & Yurich, K. K. L. (2012). *The precision teaching book*. Lemont, PA: Greatness Achieved.

Kunnskapsdepartementet. (2013). *Kunnskapsløftet [The Knowledge Promotion]*. Retrieved from http://www.regjeringen.no/en/dep/kd/Selected-topics/compulsory-education/Knowledge-Promotion.html?id=1411

Kunskapsdepartementet. (2015). *Realfagstrategi - Tett på realfag [The sciencie studies strategy - close to science studies]*. Retrieved from https://www.regjeringen.no/no/dokumenter/tett-pa-realfag/id2435042/

Lin, F.-Y. & Kubina, R. Jr. (2005). A preliminary investigation of the relationship between fluency and application for multiplication. *Journal of Behavioral Education, 14*, 73–87. doi: 10.1007/s10864-005-2703-z

Murnane, R. J., Willett, J. B. & Levy, F. (1995). The growing importance of cognitive skills in wage determination. *Review of Economics and Statistics, 77*(2)251–266. doi: 10.2307/2109863

Olsen, S. C., Daly III, E. J., Anderson, M., Turner, A. & LeClair, C. (2007). Assessing student response to intervention. In S. R. Jimerson, M. K. Burns & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention. The science and practice of assessment and intervention*. (pp. 117–129). New York NY: Springer Science+Business Media.

Shapiro, E. S. (2011). *Academic skills problems (4ed.).* New York, NY: The Guilford Press.

Strømgren, B., Berg-Mortensen, C. & Tangen, L. (2014). The use of precision teaching to teach basic math facts. *European Journal of Behavior Analysis, 15*, 225–240. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/15021149.2014.11434723

Utdanningsdirektoratet. (2015a). *Tall fra grunnskolens informasjonssystem (GSI) 2014/15.* Retrieved from http://www.udir.no/globalassets/upload/statistikk/gsi/grunnskole-gsi-notat-2014-15.pdf

Utdanningsdirektoratet. (2015b). *Skoleporten*. Retrieved from https://skoleporten.udir.no/rapportvisning/grunnskole/laeringsresultater/eksamenskarakterer/nasjonalt?rapportid=2&diagraminstansid=1&enhetsid=00&vurderingsomrade=11&underomrade=21&skoletype=0&oversikttypeid=0&fordeling=2&skoletypemenuid=0&sammenstilling=1&kanviseprikking=0&barevisoffentligedata=True&indikator=381&diagramtype=3

Utdanningsdirektoratet. (2016). *Nasjonale prøver*. Retrieved from http://www.udir.no/Vurdering/Nasjonale-prover/#Hva-er-nasjonale-prover

VanDerHeyden, A. M., Witt, J. C. & Gilbertson, D. (2007). A multi-year evaluation of a response to intervention (RtI) model on education of children for special education. *Journal of School Education, 45*, 225–256. doi:10.1016/j.jsp.2006.11.004

VanDerHeyden, A. M., & Burns, M. K. (2008). Examination of the utility of various measures of mathematics proficiency. *Assessment for Effective Intervention, 33*(4)215–224. doi: 10.1177/1534508407313482

VanDerHeyden, A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education, 18*, 71–91. doi: 10.1007/s10864-009-9081-x

VanDerHeyden, A. M., & Burns, M. K. (2010). *Essentials of Response to Intervention.* Hoboken NJ: John Wiley & Sons Inc.

Wong, M. & Evans, D. (2007). Improving basic multiplication fact recall for primary school students. *Mathematics Education Research Journal, 19*, 89–106. doi:10.1007/bf03217451

Wright, J. (2013). *How to: Teach students to change behaviors through self-monitoring*. Retrieved from http://www.interventioncentral.org/sites/default/files/pdfs/pdfs_blog/self_management_self_monitoring.pdf

Examining the Relationship Between Component and Composite Frequency and Natural Growth in Basic Multiplication: A Systematic Replication of Lin and Kubina

Cathrine Berg-Mortensen

and

Linn Tangen

Oslo and Akershus University College of Applied Sciences

Distributions of tasks when writing article two.

The current article has been a collaboration from the beginning to the end. Both authors have contacted the participating schools, both have registered testpackets following the data collections, and analyzed the data.

In the writing process Cathrine has mainly written the introduction and discussion sections, while Linn has mainly written the method and results sections. But we both have contributed in the progress of each others written sections, and the progression of the latest edition, which is the current article.

Abstract

Mathematics is a universal problem, in Norway the most pressing concern is the evident decline in mathematic proficiency in high school, and the increased number of students in need of special education. Assessment of student proficiency level, and possible learning disabilities, are traditionally done through norm- and criterion-referenced achievement tests. These tests will rarely show what the potential learning disabilities are, nor will they point to interventions likely to be effective. In order to implement effective interventions there is a need for further research on natural frequency growth in the transition from component to composite skills, and over grade levels. Findings from these studies could establish frequency aims that could possibly predict future student outcome, and generate more effective interventions, that could enhance student proficiency in mathematics.

The current study has examined the relationship between component and composite frequency and natural growth in basic multiplication in the grade levels 5–9, and is a systematic replication of Lin and Kubina (2005) with variations; the inclusion of multiple grade levels (5th–9th), a second screening, and a simple linear regression analysis to incorporate prediction of direction of the variables. Findings from the current study shows that component frequency predicts composite frequency and as such reinforce the findings from Lin and Kubina (2005), and also support the notion that basic skills are prerequisites of more complex skills.

Keywords: frequency, Curriculum Based Measurement, Precision Teaching, Response to Intervention, component and composite performance

International tests such as *Program for International Student Assessment* (PISA) and *Trends in Mathematics and Science Studies* (TIMSS) shows that mathematics is a universal problem. In Norway, the most pressing concern is the evident decline in mathematic proficiency in high school (Grønmo et al., 2012; Kjærnsli & Olsen, 2013). It is also shown that the need for special education is growing with grade level. In the school year 2014/2015 students in need for special education increased from 3.6% in the first grade to 11.1% in the 10^{th} grade (Utdanningsdirektoratet, 2015a). At the same time 14.5% of the Norwegian 10^{th} graders failed their final exam in mathematics (Utdanningsdirektoratet, 2015b).

The currents state of mathematical proficiency has been an impetus for an increased national debate and effort aiming to find more effective interventions (Kunnskapsdepartementet, 2013/2015). This debate has mainly revolved around structural interventions such as, increased number of teachers and adults in schools, and demanding more specialized training of teachers in regard to core subjects (Kunnskapsdepartementet, 2013/2015). Structural interventions like these are all are ranged to have a low educational effect size; Cohen's d = .40 or less, if not combined with other empirically supported interventions (Hattie, 2009). The assessment of student proficiency level, and possible learning disabilities, is traditionally done through norm-and criterion-referenced achievement tests (Shapiro, 2011; VanDerHeyden & Burns, 2010). Norm-referenced achievement tests often refer to what is an expected level of proficiency at a certain age or grade level. The results will often show a narrow part of the total picture and primarily depict whether or not the student is at the expected proficiency level. These tests will rarely show what the potential learning disability is, nor will they point to interventions that would likely be effective, i.e., the "diagnostic value" may be questionable (Shapiro, 2011).

In Norway, *National tests* are implemented in order to screen students at the start of the 5^{th}, 8^{th} and 9^{th} grade level. The main purpose of *National tests* is to assess students' basic skills in reading, mathematics and parts of the English course, and thereby be able to assure that the students have reached the expected level of academic proficiency, within the given subjects. *The National tests* results are further expected to guide the teachers in their instructional methods and the districts are also expected to use the results to assess the quality of the general education (Utdanningsdirektoratet, 2016). *Supportive-learning tests* have also been developed as an aim to assess student proficiency in mathematics in the grade levels 5–7 and 8–10. However, the schools themselves decide how often these tests are to be administered (Kunnskapsdepartementet, 2015). Both *National tests* and *Supportive learning tests* are examples of criterion-referenced tests and have the advantages of assessing student performance towards a standardized level based on the Norwegian standard curricula. However, these assessments are administered at a low frequency and, as such, show some limitations in detecting academic problems at an earlier time (Shapiro, 2011).

In spite of previously and current effort by the Norwegian government, student proficiency in mathematics have not improved since 2003 according to the international TIMSS and PISA tests (Grønmo et al., 2012; Kjærnsli & Olsen, 2013). Thus the need for more specific and effective screening procedures with improved "diagnostic value" seems to be desirable (VanDerHeyden & Burns 2010; Shapiro, 2011; Johnson & Street, 2013).

Effective Screening Procedures

More effective and brief assessments procedures were developed as an alternative approach to norm- and criterion-referenced tests, and aimed at being time efficient and highly predicable in regards to students' future academic outcome.

A specific universal screening procedure is Curriculum based measurements (CBM) (VanDerHeyden & Burns, 2010; Shapiro, 2011). Deno (2003) writes that CBM originated as a tool for assessing the basic skill proficiency of students with learning disabilities. The generic CBM procedure is direct measurement of student proficiency with stimulus materials used by the teacher in the classroom, in order to assess the academic growth and outcome over time. In more recent years there has been an increase in the use of CBM as a standardized procedure for assessing students at risk for academic failure, in the four basic skill areas; reading (CBM-R), mathematics (CBM-M), written expression (CBM-W), and spelling (CBM-S). Student performance is repeatedly measured through the academic year, and each CBM procedure usually lasts from 1–3 minutes. Response frequency is scored by counting the number of correct and incorrect responses (VanDerHeyden & Burns 2009/2010; Shapiro, 2011).

CBM-M procedures are used to assess both basic and broader mathematical proficiency. When used to assess basic skills the duration of CBM-M is usually 2–8 minutes. Standard measurement unit is correct and incorrect responses with regard to the corresponding expected grade, goal, or instructional level during the timing interval (Christ et al., 2008). CBM-M is often referred to as a valid and reliable set of measurements, VanDerHeyden and Burns (2008) found that CBM-M as a measurement procedure applied to investigate fluency is most reliable to predict academic outcome in group-based administration of mathematics. However, CBM-M has been less researched than CBM-R and CBM-W, which makes it difficult to determine the growth or possibly the positive outcome for students in mathematics (Christ et al., 2008; VanDerHeyden & Burns, 2009). Further, it is important to notice that brief assessments of basic skills, are a measurement of one type of responses at a given point in time, and that the results from such assessments are to guide further interventions within the given domain (e.g. reading, mathematics etc.). The current recommendation is to use CBM-M as an assessment tool for specific subset skills of mathematics (Christ et al., 2008).

Another approach to brief and repeated universal screening procedures is *Brief Experimental Analysis* (BEA). BEA can be used as a tool for assessing where intervention is needed and further increase the probability for positive learning outcomes for students (VanDerHeyden & Burns, 2009).

Precision Teaching (PT) has also established effective screening procedures through the development of the Standard Celeration Charts. These charts monitor skill development at a frequent basis (day–day/session–session) and as such, offers crucial information on how a frequency changes as a curricula or intervention advances, and whether or not the current curricula or intervention is effective upon student learning (Johnson & Street, 2013; Kubina & Yurich, 2012).

PT, BEA and CBM have amongst some tried to investigate if a given response frequency can predict academic growth, from basic subset skills to more advanced mathematical skills. Frequency aims that could possible predict academic growth are in PT termed fluency, and refers to the ability to master a skill with accuracy plus speed (Binder, 1996) that produces the following outcomes, abbreviated as MESsAGe; (1) Maintenance, (2) Endurance, (3) Stability, (4) Application, and (5) Generativity. Given that a response frequency is fluent it also has to be (1) maintained in the future, (2) over longer period of times, (3) in other stimulus environments, (4) applied in a real world context, (5) and emerge in to more complex behavior repertoires (Johnson & Street, 2013). The concept of fluency is well described in most PT literature, but there is still a lack of evidence on what level of frequency that is necessary to reliably predict fluency (Binder, 1996; Haughton, 1972).

Haughton (1972) wrote that the best information one had at the time showed that writing single digits accurately at a frequency of 80 or above per minute, or computing 40–50 problems per minute was a good indicator for academic growth in mathematics. A frequency level at 30 or fewer math fact completions per minute would hence indicate deceleration as the curriculum advanced. Further studies have found the same frequency aim to be a predictor for future mathematical performance, but there is still a need for more research to establish a frequency aim that can predict positive learning outcomes for students in the total domain of mathematics (VanDerHeyden & Burns, 2009; Christ et al., 2008; Johnson & Street, 2013).

VanDerHeyden and Burns (2009) examined whether it is possible to develop cut-off scores in mathematics to identify students at risk for declining, and further if it is possible to identify a fluency criterion that can predict retention through BEA. Results indicated that the students mastering basic skills were more likely to reach the fluency criterion than the students that did not master basic skills. As a conclusive statement VanDerHeyden and Burns (2009) refers to that there is a need for developing cut-off scores in mathematics that can be used; to identify students at risk for academic failure, and as useful criteria to evaluate efficacy of different interventions, and further shows potential academic outcomes for the students in general. CBM-M shows the potential to be effective in assessing student proficiency but faces certain challenges (Christ et al., 2008). The main challenge is whether or not a certain level of frequency can be established as a potential cut-off score, or fluency aim that can reliably predict future academic outcome for the students (VanDerHeyden & Burns, 2009).

The notion of teaching academic skills in any area is that component skills are a prerequisite for more complex/composite skills, thus they are a necessity for the individual academic growth (Johnson & Street, 2013; Binder 1996; Kubina, Kostewich & Lin, 2009). The taxonomy of learning can yield information on how heightened competence can be achieved. Kubina et al. (2009) refers to the taxonomy of learning approach by Tiemann and Markle that comprises of four levels; (1) psychomotor, (2) simple cognitive, (3) complex cognitive and (4) emotional. In mathematics, the levels most susceptible to instruction are simple cognitive and complex cognitive, and the acquisition of knowledge from simple cognitive to complex cognitive is derived through the six stages of learning (Kubina et al., 2009). The six stages of learning are an illustration of pre-fluency skills (entry and acquisition) as well as the different fluency outcomes (Johnson & Street, 2013).

Shapiro (2011) writes that there is a compelling need for a methodology that can more accurately and directly assess basic academic skills, that can produce positive learning outcomes for the student.

Response to Intervention

A framework proven to be effective in improving academic outcomes for students is *Response to Intervention* (RtI) (Johnson & Street, 2013). RtI is founded on multiple methods and procedures developed in applied behavioral analysis, such as PT, CBM and BEA among others (VanDerHeyden & Burns, 2010; Johnson & Street, 2013).

RtI contains different strategies for assuring the quality and expected learning outcomes for the students. Johnson and Street (2013) mentions (1) early identification, (2) early interventions, (3) high-quality, empirical supported core curriculum and interventions, increasingly intensive interventions, (4) continuous and increasingly frequent progress monitoring and (5) data-based decision making as the core components of RtI. The RtI model is visualized, as a triangle with a wide bottom and narrow top, which illustrates the service delivery needs of the different tiers. The bottom level (Tier 1) refers to all students, where an empirical supported core curriculum is in effect and with the main purpose of having a preventive and proactive effect, 80–90% of students are found in Tier 1. Tier 2 illustrates students with similar error patterns who are in need for more specialized intervention, the interventions are usually group based and focuses on being efficient and create rapid responding. In Tier 3, students with individual learning disabilities with the need for more individual specific interventions are found (VanDerHeyden & Burns, 2010; Johnson & Street, 2013; Shapiro, 2011).

The triangle model is not a new phenomenon, in the medical domain, this have been an efficient way of screening patients for correct treatment. Tier 1 would refer to the general medical service, which is available for all to prevent potential disease outbreaks; an example is the vaccination program for newborns. Tier 2 and 3 would refer to groups or individuals in need for more specialized treatment to prevent diseases. Furthermore, the triangle is a known model of screening students for behavior problems, termed *Positive Behavior Interventions and Support* (PBIS). PBIS and RtI share many common features mainly that they both are preventive models (Johnson & Street, 2013).

Brief universal assessments, such as CBM, and progress monitoring are two of the five components of the RtI framework, and comprise the foundation for further decisions regarding potential effective interventions that could increase student learning. PT has also been proposed as a complementary measurement procedure within a RtI framework for assessing student proficiency at the different tiers, and also as an effective intervention for students in Tier 2 and 3 (Kubina & Yurich, 2012; Johnson & Street, 2013).

Universal screening procedures and progress monitoring can generate more accurate and reliable decisions for future interventions. However, for a RtI framework to prove effective in preventing student decelerating and enhancing student proficiency in mathematics, there is a compelling need for establishing frequency aims that could predict future mathematical growth with and without effective interventions.

Component and Composite Performance

Lin and Kubina (2005) conducted a study with 156 5th grade level students examining the relationship between the concepts of accuracy and fluency[1] in component (single-digit) and composite (multi-digit) multiplication. Their research questions were: “(1) How well will the participating 5th grade students perform basic multiplication facts and multi-digit multiplication problems in terms of accuracy and fluency,” and (2) “What is the relationship between basic multiplication fact performance and multi-digit problem completion" (Lin & Kubina, 2005, p.76).

The Lin and Kubina (2005) results for accuracy showed a mean of 98.12% for the component skill (single-digit) and a mean of 88.40% for the composite skill (multiple-digit). There was a medium but significant correlation (r = .371, $p < .01$[two-tailed]) between component and composite accuracy. Results for frequency showed a mean frequency of 57.01 for the component skill, and only 13.1% scored above the frequency aim of 80 correct digits per minute. For the component skill the mean frequency was 88.40, and only 3.2% scored above the frequency aim of 40 correct digits per minute. There was, however a strong and significant correlation between component and composite frequency ($r = .745$, $p < .01$[two-tailed]).

The correlation between component and composite frequency reported by Lin and Kubina (2005) may indicate that frequency building in component skills can improve frequency in composite skills. There is an uncertainty about this however, as the correlation measure used does not indicate the direction of the correlation. A more refined measure such as simple linear regression analysis (Yockey, 2011) could more accurately indicate whether component frequency levels predicts composite frequency levels. This kind of knowledge can lead to effective teaching by for instance, assuring that student's component skills are mastered to a certain degree or criteria before embarking upon more complex/composite skills.

The current study is a systematic replication of Lin and Kubina (2005). The systematic variations are: (1) inclusion of 6th–9th grade levels in addition to the original 5th grade level, (2) inclusion of a second screening for all grade levels (5th–9th), and (3) a simple linear regression statistical procedure to investigate whether component frequency and accuracy predicts composite frequency and accuracy.

The research questions for the current study were partly in accordance with Lin and Kubina (2005): (1) “How well will the participating 5th–9th grade students perform basic multiplication facts and multi-digit multiplication problems in terms of accuracy and fluency,” (p. 76). For the current study, the second research question is reformulated, and reads: (2) “To what degree do basic multiplication fact performance predict multi-digit problem completion.” An additional research questions was: (3) “What is the natural growth in numeral writing frequency, single and multi-digit frequency and accuracy for the different grade levels.”

Method

Participants

Three middle schools (5th to 7th grade level) and four high schools (8th to 9th grade level), all located in the Romerike region in Norway, participated in the current study. During the first data collection (T1) a total of 779 students were included: 66 from the 5th grade level, 60 from the 6th grade level, 63 from the 7th grade level, 243 from the 8th grade level and 347 from the 9th grade level. During the second data collection (T2) a total of 626 students were included, and the corresponding number of participants for each grade level were 22, 50, 37, 197, and 320 accordingly.

In order to recruit schools, the experimenters contacted principals in the designated school area, explained the study and asked if they were interested in participating. Seven schools accepted. All students attending class the day the experimenters arrived had the opportunity to participate. However, the students them self could refrain from participating. Since all the students’ answered test packets were anonymous (i.e., no traceable personal information) written parental consent were not needed.

Materials

The same test packets that Lin and Kubina (2005) constructed were used in the current study (Appendix), with approval. The test packet consisted of three different sections. Section one measured the students' number writing skills, a screening measure, where the students were told to write the Arabic numerals from 0 to 9 repeatedly on lined notebook paper. Section two presented the students with 156 random single-digit multiplication problems (e.g., 8 x 5) in order to assess their Component skill performance. Section three presented the students with 63 random multi-digit multiplication problems (e.g., 986 x 83) in order to assess their Composite skill performance. The iPhone 6 Clock application's Timer feature was used to start and end the 1-minute assessment intervals. Pens were preferred when answering the test packets and the students used their own.

Dependent Variables/Measurement Units

As in the Lin and Kubina (2005) study, the measurement unit of the current study was correctly written digits per minute. The total correct digits generated two variables—frequency and accuracy. Each correctly written digit was counted and included in the student's total frequency and accuracy score. The frequency variable was measured as the number of correctly written digits per minute, and the accuracy variable was measured by dividing correctly written digits on the total digits in each algorithm, which gave a percentage number correct digits. This yielded four variables to analyze: Component skill frequency, Component skill accuracy, Composite skill frequency and Composite skill accuracy. Also, a screening variable, Number writing frequency, was calculated.

The pre-constructed key described in Lin and Kubina (2005) was used. For example, the multiplication problem 8 x 5 = 40 gives two correct digits (frequency) and 100% accuracy. If the student made a mistake and wrote 41 as the final answer then it would be counted one correct digit (i.e., the 4 in the tens column is correct, but the 1 in the ones column is not), and 50% accuracy. In the case of 986 x 83, this includes an algorithm of 2958 + 78880 = 81838, which gives the total of 14 correct written digits (frequency) and 100% accuracy. But if a student only wrote 2958 then four correct digits was counted toward the students' total frequency, but the accuracy would still be 100% because all written digits are correct. However, if the student has calculated the whole algorithm (i.e. final answer: 2958 + 78880 = 81838) with all digits correctly placed, the accuracy will also be 100%.

The frequency aims in the current study equals those used by Lin and Kubina (2005). Component skill (single-digit) frequency aim was set at 80 correctly written digits per minute (the usual aim is 80–120 according to Lin and Kubina (2005)), and Composite skill (multiple-digit) frequency aim was set at 40 correctly written digits per minute (the usual aim is 40–60 according to Lin and Kubina (2005)). Both the Component and Composite accuracy aim was set at 100% correct.

Interscorer Agreement and Reliability

Both interscore agreement and Intraclass Correlation Coefficient will be assessed as there are some differences about the usability of the different reliability measures (Gast & Leford, 2014; Hallgren, 2012).

Interscore agreement was assessed for every tenth test packet, all measure sections were counted and correct and total digits were registered. The T1 mean interscorer agreement was 100% for all measures; Numeral writing, Component frequency, Component accuracy, Composite frequency and Composite accuracy respectively. The T2 mean interscorer agreement was 100% for the measures Numeral writing, Component frequency and Component accuracy respectively. For both Composite frequency and Composite Accuracy the mean interscorer agreement was 99.9% (Min. 91%, Max. 100%).

Reliability was also measured for every tenth test packet for the outcome variables Composite frequency and Composite Accuracy. The Intraclass Correlation Coefficient by means of a two-way mixed model with average measure, expressed as ICC (3,2), yielding a reliability measure expressed as Cronbach's alpha (α) with 95% Confidence Intervals reported in the format of [LL, UL]. Results for T2 showed that $\alpha = 1.00$ for Composite frequency, and Composite accuracy respectively.

Procedure

The procedure was a replication of the procedure used by Lin and Kubina (2005). All the tests were completed in the students' regular classroom at their school. The students were sitting at their regular desks and the experimenters presented themselves and gave a short description of the study and what was expected of the students. Test packets were handed out to all students with instruction of not turning them around until the experimenters said so. The experimenters informed that they should answer each of the three test sections in one-minute as quickly and accurately as possible. The experimenters used the same scripted message as Lin and Kubina (2005), only translated to Norwegian.

> You are going to do three short tests. Each test lasts one-minute. I will use this timer to set one-minute. For each test, when I say "ready set go," please start the test. When the timer starts to beep, please stop and put your pencil on the table (Lin & Kubina, 2005, p. 77).

The Norwegian translated message:

> Nå skal dere gjennomføre tre korte tester. Hver test har en varighet på ett-minutt. Jeg kommer til å bruke denne iPhone og sette på ett-minutt. For hver test, når jeg sier ”klar, ferdig, gå" så kan dere starte testen. Når klokka starter å pipe, vær snill å stoppe og legg ned blyanten på pulten.

Each section of the test had additional instructions. In section one of the test the students were told to repeatedly write 0–9 on lined paper. This was demonstrated by the experimenter using her hand writing from left to right in the air, simulating an answer. In both sections two (single-digit) and three (multiple-digit) of the test the students were told to cross out their answer if they did it wrong, and just to skip a problem they did not know the answer to. They were further informed that time is of the essence and that they should not waist the time to erase wrong answers. The sections two and three both included two pages each, and the students were told to start on page one and if they finished then go to page two. All students were told to write out their full answer, specifically on section three they were requested to write the total algorithm in addition to the final answer.

Questions from students were answered before the test was started, and all the students were informed that the tests contained a lot of problems, and they were not able to finish them all in one minute. When all three sections of the test were finished the students were told to write a code on the first page. The code was the first three letters in the students' mother's first name, then the first three letters in their own first name, then the first three letters of their father's first name. The code made it possible to compare students' test from both data collections (T1 and T2) and still preserve anonymity.

The first data collection was conducted in end of January 2015, and the second data collection was conducted in the beginning of June 2015. The days between the first and second data collection ranged from 126 to 143 days for the 7 schools. The same procedure was used on both data collections.

Statistical Analysis Plan

In order to answer the first research question, on how well will participating students perform single-digit multiplication problems and multi-digit multiplication problems in terms of accuracy and frequency aim, descriptive statistics were used to analyze the data collected (Tables 1, 2, and 3).

The second research question was examined using inferential statistics, more specifically a simple linear regression analysis, in order to examine to what degree a score on one (independent) variable predict a score on another (dependent) variable (Table 4). This analysis gives information about the direction of the data, not merely the correlation, and yields a predicting value termed beta (β). Beta is the average amount by which the dependent variable increases when the independent variable increases one standard deviation and other variables are held constant (Yockey, 2011), an expression of the predictive importance of the independent variable. Beta is equal to Pearsons r in simple linear regression. The effect size expression in simple linear regression is R^2 with the values.01, .09 and .24 corresponding to small, medium and large effect sizes respectively. R^2 is also an expression for «explained variance», how many percent of the variance in the dependent variable that is explained by the independent variable. Section one of the test packet measured the students' writing speed, and it was a way to find out which student who might have slow handwriting skills. If a student's Numeral writing performance was lower than the Composite frequency aim, 40 digits per minute, the student was excluded from the regression analysis.

In addition to the research questions that Lin and Kubina (2005) studied, the current study also looked at the students' growth from data collection one (T1) to data collection two (T2) in the different grade levels. For those students who participated on both occasions (T1 and T2), a repeated measures t-test was calculated for all outcome variables (Table 5) for each grade level. The Cohen's d effect size was calculated with the formula Mean difference/Standard deviation of the difference scores, and d conventions for small, medium and large effect sizes are .20, .50, and .80 accordingly (Yockey, 2011).

Results

779 completed test packets from data collection on (T1) were analyzed. Due to space limitations only 5th grade results will be written out, this because it is the comparison with Lin and Kubina (2005) results that is most relevant to examine. Results for the other grade levels are depicted in Tables 1, 2 and 3. All frequencies represent correct digits per minute.

Mean, Min., and Max. Comparisons

Results for all grade levels are depicted in Table 1. Fifth grade students had a Mean numeral writing frequency of 60.58 (Min. 25, Max. 92), a Mean Component frequency of 17.53 (min. 0, max. 45), and a Mean Composite frequency of 2.03 (Min. 0, Max. 8). Mean Component accuracy was 68.21% (Min. 0%, Max 100%), and Mean Composite accuracy 57.97% (Min 0%, Max. 100%).

Fifth grade students performed poorer than the other grade levels on both Component and Composite frequency, but performed better or equal to the other grade levels on Component and Composite accuracy.

Comparison with Performance Aims

Results for all grade levels are depicted in Table 2. 88% of 5th grade level student performed at or above a frequency of 40 on Numeral writing, 12% did not. None of the 5th grade level students performed at the frequency aim (≥ 80) on Component frequency, whereas 62% performed at at the aim (100%) on Component accuracy. None of the 5th grade level students performed at the frequency aim (≥ 40) at Composite frequency, but 43% performed at the aim (100%) on Composite accuracy.

In total only three students performed at the Component skill frequency aim (≥ 80), two from the 8th grade level and one from the 9th grade level. Furthermore, the Composite skill frequency aim (≥ 40) was not reached by any of the students.

Variability among Accuracy and Fluency

Results for all grade levels are depicted in Table 3. 62% of 5th grade level students had 100% Component accuracy, whereas 38% had below 100%. For those with 100% Component accuracy the Mean Component frequency was 20.32 (SD = 9.32) (Min. 4, Max. 45). For those with less than 100% Component accuracy the Mean Component frequency was 12.96 (SD = 8.42) (Min. 0, Max. 37).

44% of 5th grade level students had 100% Composite accuracy, whereas 56% had below 100%. For those with 100% Composite accuracy the Mean Composite frequency was 2.66 (SD = .90) (Min.1, Max. 4). For those with less than 100% Composite accuracy the Mean Composite frequency was 1.54 (SD = 2.05) (Min. 0, Max. 8).

Regression Analysis Between Component and Composite Frequency and Accuracy

Results for all grade levels are depicted in Table 4. Eight students from the 5th grade level, two from 6th grade level, four students from 7th grade level, 12 from 8th grade level, and nine from 9th grade level performed below the frequency aim of 40 written digits per minute on numeral writing (see Table 2). These 35 students' test packets were excluded from the simple linear regression analysis.

When examining if Component frequency predicts Composite frequency in 5th grade level students, there were no significant predicting value; $\beta = .136$, $p > .05$ and Component frequency accounts for only 1.8% ($R^2 = .018$) of the variance of Composite frequency. Furthermore, for 5th grade level students Component frequency predicted lower Composite accuracy, $\beta = -.284$, $p < .05$, with a small ES ($R^2 = .018$). Regarding Component accuracy for 5th grade level students, this variable did not predict Composite frequency ($\beta = .104$, $p > .05$, $R^2 = .011$) or Composite accuracy ($\beta = -.07$, $p > .05$, $R^2 = .005$).

Further results show no significant relationship examining the variables Component frequency and how it predicts Composite frequency, but 6th, 7th, 8th and 9th grade levels show medium ES.

For most of the participating students the analysis shows no effect at all, the only variable showing an effect is the Component frequency predicting Composite frequency. For the other three variables in Table 4, Component frequency predicting Composite accuracy, Component accuracy predicting Composite frequency, and Component accuracy predicting

Composite accuracy, the predicting levels are to low and not significant for any of the grade levels.

Repeated Measures t-test for Numeral Writing, Component Frequency, Component Accuracy, Composite Frequency and Composite Accuracy.

Results for all grade levels are depicted in Table 5. There were 17 5th grade level students who completed both data collections. Numeral writing showed a significant growth from T1 (M = 57.88, SD = 14.80) to T2 (M = 65.00, SD = 15.13), $p < .05$, with a medium ES (d = .69). Component frequency also showed a significant growth from T1 (M = 11.65, SD = 7.049) to T2 (M = 19.41, SD = 9.23), $p < .01$, and the ES is large (d = 1.11). Component accuracy showed a significant change from T1 (M = 86.77, SD = 18.54) to T2 (M = 96.28, SD = 7.17), $p < .05$, and the ES was medium (d = .53). Composite frequency and Composite accuracy showed no significant change from T1 to T2.

Overall, the significant changes from T1 to T2 are found in Numeral writing and Component frequency. A few grade levels achieve significant results in Component accuracy. For Composite frequency only 9th grade level showed a significant improvement from T1 to T2, and for Composite accuracy none did.

Discussion

The current study aimed to investigate (1) "How well will the participating 5th–9th grade students perform basic multiplication facts and multi-digit multiplication problems in terms of accuracy and fluency", (2) "To what degree do basic multiplication fact performance predict multi-digit problem completion" and (3) "What is the natural growth in numeral writing frequency, single and multi-digit frequency and accuracy for the different grade levels". A total of 779 test packets from T1 and 626 test packets from T2 were analyzed. Due to space limitations the results from the 5th grade will be emphasized, the remaining grade levels will mainly be discussed in regards to the third research question.

The data from the current study illustrates that overall most students reached the frequency aim for numeral writing, only 4.5% did not. None but three of the students (grades 5–9) reached the component frequency aim, and none reached the composite frequency aim (grades 5–9).

For component accuracy 62% of 5th graders reached the aim, and 43% in composite accuracy, the picture being similar in grades 6–9 for component accuracy but not for composite accuracy. Results indicate that component frequency predicted composite frequency with a medium effect for 6th–9th grade, but not for 5th grade. Component frequency did not predict accuracy (component or composite).

The natural growth from January to May 2015 showed an increase in mean frequency in component frequency for all grade levels with medium to large effect sizes. Only 9th grade showed a significant increase in composite frequency with a small effect size. In component accuracy only 5th and 9th grade showed a significant increase with a medium and small effect sizes. None of the grade levels had a significant increase in composite accuracy.

General Student Performance

Numeral Writing

Results from the current study illustrates that the majority of students reached the frequency aim of writing single-digit numbers from 0–9 at a frequency at 40 digits or more per minute. The results showed however that the participating 5th grade students in the current study had a lower frequency in numeral writing than did the 5th grade students in the Lin and Kubina (2005) study. Overall the Norwegian 5th graders held a lower minimum and maximum frequency, which gives a lower mean frequency than the American 5th graders. In general, none of the participating grade levels in the current study reached the same mean frequency as the American 5th graders. These results indicate a trend of lower frequency through the rest of the test packages among the Norwegian participants in comparison to the Lin and Kubina

(2005) study. A slower frequency among the Norwegian participants in numeral writing may indicate overall slower performance through the test-packets than the American 5th graders in component and composite skills.

Potential causes for the difference in numeral writing is difficult to determine, but there is a possibility that the curricula, teaching strategies, and emphasis on tool skills, such as handwriting, differs among the two countries and that could serve as a potential variable for the numeral frequency differences. Differences between the countries and the time passed since the Lin and Kubina (2005) study can indicate that students in the current study more often use keyboards for writing and calculators in mathematics, thus, the time spent practicing tool skills is reduced.

Component Skill Performance

The results in component skills in the current study illustrates that the majority of students in all grade levels reached the accuracy aim, but none except 3 students reached the frequency aim in component skills. The Norwegian 5th graders showed none students reaching the frequency aim, which illustrates a difference from the American 5th graders, where 22 students reached the component frequency aim. Which further strengthens the assumption that Norwegian students hold a lower frequency through the test-packets than the American 5th graders did.

When the accuracy aim in component skills was held consistent at 100%, the results show a greater variance in the 5th grade component skills, than the American 5th graders. The 5th grade students in the current study who scored below accuracy aim, also illustrates a lower frequency level. Which indicate that higher accuracy levels to a degree corresponds with higher frequency levels. However, they still performed below the frequency aim at 80 correct digits per minute. The variance among accuracy and fluency in the 6th grade is more equal to the level of variance displayed in the Lin and Kubina (2005) study. However, the general trend shows a larger variance between component frequency and accuracy scores through the grade levels. The overall low frequency in component skill could indicate poorer results in composite skills. Given that high frequency scores with accuracy would yield further celeration as the curricula advances, and that low frequency scores with accuracy would yield a deceleration as the curricula advances (Binder, 1996; Haughton, 1972).

Reasons for the overall low frequency scores may be due to a lack of proficiency in tool skills, but may also reflect an acquisition level in component skills even in the higher grade levels. This may be a function of too rapid progress through the curriculum, i.e., the students do not get a chance to move through the different stages of learning before a new skill is introduced (Kubina et al., 2009).

Composite Skill Performance

In composite skill performance 43% of the participating 5th graders reached the accuracy aim, but none reached the frequency aim of 40 correct digits per minute. In comparison; 3.2% of the 5th grade students reached the frequency aim, and 48.1% reached the accuracy aim in the Lin and Kubina (2005) study. In comparison with component skills, the assumed trend of deceleration in frequency and accuracy seems to hold true.

None of the participating students in the current study reached the frequency aim, and the majority of students did not reach the accuracy aim. The variability among frequency and accuracy is greater in composite skills than component skills, and in general, except the 5th grade, the amount of students scoring at the accuracy aim holds a higher frequency than students that score below the accuracy aim. Which is similar to the trend in component skills.

A factor worth noting in regards to the composite accuracy results is how correct digits were scored towards the accuracy percentage. The study by Lin and Kubina (2005) counted every correct digit in correct place as correct. As the current study is a systematic replication, the identical scoring method were applied, but was during the study discovered to be an unclear scoring method. Unfortunately, this was discovered too late. A more precise method could be; counting percentage of correctly placed digits for the whole algorithm. If the algorithm in total contains 13 digits, then 13 are the required number to achieve 100% correct (Kubina et al., 2009, p.23).

Component and Composite Frequency and Accuracy

In the current study results illustrates that component frequency predicted composite frequency with a medium effect, but not composite accuracy. Lin and Kubina (2005) found similar results; a strong correlation between component and composite frequency, but not between component frequency and composite accuracy. The correlation between component and composite frequency is consistent with previous research on fluency, which shows that frequency levels (high or low) in component skills correlates with frequency levels (high or low) in composite skills (Haughton, 1972; Lin & Kubina, 2005; Johnson & Street, 2013; VanDerHeyden & Burns, 2009).

Explanations for why component frequency did not predict composite accuracy could be a lack of proficiency in other component skills needed to compute the composite math facts other than single-digit multiplication, such as addition. Other factors could be that the students used the timing interval searching for math facts they would be able to compute.

Component accuracy did not predict composite frequency or accuracy, which is similar to the results from Lin and Kubina (2005). Explanation for this could be a lack of proficiency in tool skills such as place value, or component skills such as addition and single-digit multiplication (Johnson & Street, 2013). Such tool and component skills are needed for completing the composite math facts. Proficiency in tool skills are in general crucial prerequisites for more advanced mathematical problems. Which could be a possible explanation for the low frequency and accuracy score attained by the students in component and composite skills in the current study.

Component and Composite skill performance in the current study shows that students in the 5th grade hold a lower frequency than the other grade levels, but scores higher or equal to the other grade levels in regards to accuracy, both in component and composite skills. The current findings indicate that the students are more focused on getting the mathematical problems correct than they are sensitive to the time spent on each problem. Working slow may at times be functional, but in the case of student performance, there is in most cases a timeframe to be sensitive to. Exams and tests usually have a time limit, and given that one works slowly it may indicate an incomplete assignment, which could reflect the grade in mathematics. Which could impact further educational decisions, and in a long time scale impact national wealth and growth (Hanushek & Kimko, 2000).

Natural Growth

Results from T1 show that the mean frequency in numeral writing and component frequency increases with grade. There is a significant growth in all grade levels in component frequency. Although all grade levels illustrate growth, individual differences among grade levels in component frequency depict a drop in T1 from students in the 8th grade to 9th grade students in T2. In component accuracy only the 5th and 9th grade showed significant growth, but there is a decrease in component accuracy after the 7th grade in both T1 and T2.

In composite frequency only the 9th grade students illustrate a significant growth, but there is a drop in component frequency after 7th grade, which is illustrated in both T1 and T2. In composite accuracy none of the grade levels had significant growth, but the same trend is seen, i.e. a drop in accuracy after the 7th grade.

Overall the grade levels are all well below the frequency aims in component and composite skills, which could indicate that they might need a higher frequency to be able to maintain the required skill. The students' natural growth in the current study also illustrates that even though they hold a low frequency, they increase through the school year, but at the same time there is a drop in frequency between grade levels at certain points.

The results show that general education is effective to a degree. Hattie (2009) states that all educational effects below a benchmark at Cohens´s d = 0.40. is in need for further analysis, that is not saying that the educational method in effect is worthless, but that they need more investigation. At the same time the results correspond with international tests, such as PISA and TIMSS which report that mathematical proficiency among Norwegian 8th grade students are in decline. Given that these results only display a given point in time, makes it difficult to determine if the results is a representative picture of natural growth in mathematics. A more comprehensive approach would be to follow grade levels over time.

General Discussion and Conclusion

The current study has a couple of strengths in comparison with the Lin and Kubina (2005) study. First, multiple grade levels were included and thus offered the opportunity to examine the frequency and accuracy variables across grade levels. This made it possible to discover that low component and composite frequency is relative stable over the grade levels. This strengthens the notion that students may not attain a proficient level in tool or component skills before they are required to progress to more complex/composite skills.

Second, the current study performed two assessments in contrary to one. The first one was conducted during the first month of the spring semester, and the second one was administered in the last month of the semester. Performing two assessments made it possible to investigate natural growth in component and composite skill under the general teaching curriculum. The findings from T1 and T2 reinforce the notion that low frequency is relative stable over the grade levels, and that the students do not seem to reach the acquisition level in basic skills, by the fact that component skill improved from January to May/June but seem to decrease again in January for the nest grade level.

A third strength of the current study is the inclusion of a simple linear regression analysis. By including this statistical analysis, we could examine both the correlation and the direction of the correlation between the investigated variables, which is an expansion of the correlation analysis applied by Lin and Kubina (2005).

However, conclusion derived from the findings could be weakened by the fact that the assessments performed are snap-shot pictures from each classroom, and the question of growth over grade levels and/or months can thus only be presumed. Future research should make efforts to follow a group of students in a grade level, for instance 5th grade, over several years and perform CBM three times each school year, as proposed in the RtI framework (Johnson & Street, 2013; VanDerHeyden & Burns, 2010; Shapiro, 2011). This could make it possible to assess whether the same component and composite skills grow across the academic year, and also whether there is a decline from the end of one school year to the beginning of the next school year, in the same student group. This would make assumptions about possible frequency growth or decline and possible explanations for the academic changes, more sound.

Other potential weaknesses could be that one school from the grade levels 5–7 dropped out at T2. This is unfortunate and could serve as a weakness, even if it was uncontrollable. This may have affected the results in natural growth from T1 to T2, due to a lower number of participants in T2. However, it did not affect the analyses that were a direct replication of Lin and Kubina (2005), as that certain school participated in T1, and T2 data only refers to the third research question. Further, did the study not include a gender variable when the test packets were performed and collected. This may have provided less nuanced data and hence conclusions from the study. Since we do not have a gender variable we do not know whether there is a gender difference in skills. Further studies are recommended to include gender.

Another weakness could be the accuracy calculation applied in the current study, as it may not represent true accuracy. The accuracy calculation described by Kubina et al. (2009) might reflect a more accurate accuracy calculation, and are recommended in further studies. The lay out of the test packet (Appendix) also serve as a potential weakness, further studies should consider to arrange a different layout of the test-packet where place-value are provided in to visualizes where to write the digits for completion of the multi-digit math facts. Future research should also add new assessment packages each year in addition to some stable core packages. These additions should contain skills targeted as end of year goals/learning objectives for each grade level. This will make it possible to assess development of more complex skills and their relation to for instance critical simpler, or component, skills. Another weakness worth mentioning is that the data registration of the test-packet after screening, where done manually by the researchers on a Mac-computer which could have resulted in human error.

The current study illustrates that there is a difference in frequency level in tool, component and composite skills, from the Lin and Kubina (2005) study. A continuous lower performance among the Norwegian participants is heavily illustrated. A similar trend is also seen in international tests such as PISA (Kjærnsli & Olsen, 2013) and TIMSS (Grønmo et al., 2012) were USA shows a higher proportion of 4th graders reaching international benchmarks than Norwegian 4th graders (International Study Center, 2011). These results may reflect differences in the core-curricula and teaching strategies between the two countries.

Efforts conducted by the Norwegian government to increase student proficiency in mathematics, might not have the desired effect upon student learning, as they focus mainly on structural issues, such as more money, increase number of teachers and adults in schools and specializing teachers in core subjects (Kunnskapsdepartementet, 2013/2015). These are interventions that are ranged as having a small effect upon student learning (Hattie, 2009). A possible cause for implementing interventions that show little effect, could be the lack of adequate information regarding student performance and growth, due to infrequent and unprecise assessments.

International and national tests are examples of norm-and-criterion referenced tests. Which measure student proficiency towards a predetermined standardized level, and are administered at a low or infrequent basis. As norm-and-criterion referenced tests proposes some advantages (yield information on general and individual student performance) they have certain limitations in regards to pinpointing the specific area in the academic environment in need of change (Shapiro, 2009). This may lead to inaccurate decisions regarding potential effective interventions and thus develop ineffective teaching strategies or curricula. Brief universal screening procedures, such as CBM, are found to be more reliable to predict future student outcome. As they are aimed to assess student proficiency within a current curriculum, and as such could yield more accurate information of the academic environment and potential changes that could enhance student proficiency (VanDerHeyden & Burns, 2009/2010; Shapiro, 2011; Christ et al, 2008).

Overall the findings replicated the findings by Lin and Kubina (2005). Both studies found that component frequency predicted composite frequency, and illustrated a higher percentage of students scoring within the component/composite accuracy aim than in the component/composite frequency aim. Which strengthens the assumption that students work slow, but accurate. Results from natural growth indicate that the low frequency is stable over grade levels, but there is drop in frequency from T2-T1 at certain grade levels. This indicate that frequency building in component skills may be a necessity for further growth through more, advanced curricula. It is possible to assume that a higher frequency in combination with accuracy would be more functional as several studies have found that skill frequency in lower grades and basic skills, corresponds to acquisition in higher grades and complex/composite skills (VanDerHeyden & Burns, 2008/2009; Johnson & Street, 2013; Haugthon, 1972). Previous research has also found that frequency building in basic skills, can develop skills that are more likely to be maintained over longer period of times and as necessary key-components for more complex skills (Johnson & Street, 2013; Binder, 1996; Haughton, 1972; VanDerHeyden & Burns, 2009).

The current findings show that there might be a need for more effective interventions aimed at increasing student proficiency in basic skills. Possible solutions for establishing more specialized and effective interventions in schools, could be to apply empirically supported methods that incorporates both the taxonomy of learning and the six stages of learning (Kubina et al., 2009). A procedure found to comprise how to teach, how learning occurs and frequency building in combination with frequent progress monitoring is PT. Kubina and Yurich (2012) states that PT could be a good procedure to implement both for students with specific learning disabilities as well as whole classes, given its frequent monitoring of student progress and its focus on rapid frequency building of basic skills.

As previously discussed, too rapid advancement through the curricula may be a cause for the lack of proficiency in tool and component skills. In order to secure proficiency in one skill level before advancing to the next, there is a need for more frequent monitoring. Universal screenings procedures such as CBM is regarded as highly reliable in predicting future academic outcome in reading and writing, and show promise in assessing student proficiency in mathematics. However, there is a need for more research on natural growth frequency in mathematics, before a certain level of frequency could be established as predictable for future student academic outcome (Christ et al., 2008). By following a lower grade over grade levels and conducting CBM approximately three times a year it might be possible to establish more reliable frequency aims, and as such generate earlier and more effective interventions, such as PT (VanDerHeyden and Burns, 2009/2010; Christ et al, 2008).

Frequent monitoring of a skill development over grade levels through brief universal screening procedures has the potential to establish predictable frequency aim for future students' performance (VanDerHeyden & Burns, 2008). Establishing frequency aims could generate more accurate decisions regarding student proficiency, and the effectiveness of the current curricula and hence, lead to earlier and more effective interventions be set in place. Which again could enhance the probability of positive academic outcome for all students.

References

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163–197. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2733609/pdf/behavan00020-0021.pdf

Christ, T. J., Scullin, S., Tolbize, A. and Jiban, C. L. (2008). Implications of recent research. Curriculum-Based measurement of math computation. *Assessment for Effective Interventions, 33*(4)198–205. doi: 10.1177/1534508407313480

Deno, S. L. (2003). Developments in Curriculum-Based Measurement. *The Journal of Special Education, 37*(3)184–192. doi: 10.1177/00224669030370030801

Gast, D. L. & Leford, J. R. (Red.). (2014). *Single case research methodology. Applications in Special Education and Behavioral Sciences.* (2 ed.). London: Routledge.

Grønmo, L. S., Onstad, T., Nilsen, T., Hole, A., Aslaksen, H. & Borge, I. C. (2012). *Framgang, men langt fram. Norske elevers prestasjoner i matematikk og naturfag i TIMSS 2011 [Progress, but still a long way to go. Norwegian students' 2011 TIMSS performance in mathematics and science]*. Oslo: Akademika.

Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in quantitative methods for psychology, 8*(1), 23–34. Retrieved from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3402032/

Hanushek, E. A. & Kimko, D. D. (2000). Schooling, Labor-Force Quality, and the Growth of Nations. *American Economic Review, 90*(5)1184–1208. Retrieved from http://www2.econ.iastate.edu/classes/econ520/Huffman/documents/SchoolingLaborForceQualityandtheGrowthofNations.pdf

Hattie, J. (2009). *Visible Learning. A synthesis of over 800 meta-analyses relating to achivement*. London, United Kingdom: Routledge.

Haughton, E. (1972). Aims-growing and sharing. In J. B. Jordan & L. S. Robbins (Eds.), *Let's* *try doing something else kind of thing: Behavioral principles and the exceptional child. A report from the invisible college conference on application of behavioral principles in exceptional child education, March, 1971*. Arlington, VA: The Coucil for Exceptional Children. Retrieved from http://www.fluency.org/Haughton_Aims_1972.pdf

Johnson, K., & Street, E. M. (2013). *Response to intervention and precision teaching. Creating synergi in the classroom*. New York, NY: The Guilford Press.

Kjærnsli, M. & Olsen, R. V. (2013). Fortsatt en vei å gå. Norske elevers kompetanse i matematikk, naturfag og lesing i PISA 2012 [Still a way to go. Norwegian Students' proficiency in Reading, Mathematics and Science Literacy in the PISA Assessment 2012]. Oslo: Universitetsforlaget.

Lin, F.-Y. & Kubina, R. Jr. (2005). A preliminary investigation of the relationship between fluency and application for multiplication. *Journal of Behavioral Education, 14*, 73–87. doi: 10.1007/s10864-005-2703-z

Kubina, R. Jr., Kostewicz, D. E., & Lin, F.-Y. (2009). The Taxonomy of Learning and Behavioral Fluency. *The Journal of Precision Teaching and Celeration, 25,* 17–28. Retrieved from http://celeration.org/wp-content/uploads/2015/06/2009_JPTC_V25.01_03.pdf

Kunnskapsdepartementet. (2013). *Kunnskapsløftet* [The Knowledge Promotion]. Retrieved from http://www.regjeringen.no/en/dep/kd/ Selected-topics/compulsory-education/ Knowledge-Promotion.html?id=1411

Kunnskapsdepartementet. (2015). *Realfagstrategi - Tett på realfag [The science studies strategy – close to science studies]*. Retrieved from https://www.regjeringen.no/no/dokumenter/tett-pa-realfag/id2435042/

Shapiro, E. S. (2011). *Academic skills problems (4ed.).* New York, NY: The Guilford Press.

International Study Center. (2011). *TIMSS 2011 international results in mathematics.* Retrieved from http://timssandpirls.bc.edu/timss2011/downloads/T11_IR_M_Chapter2.pdf

Utdanningsdirektoratet. (2015a). *Tall fra grunnskolens informasjonssystem (GSI) 2014/15.* Retrieved from http://www.udir.no/globalassets/upload/statistikk/gsi/grunnskole-gsi-notat-2014-15.pdf

Utdanningsdirektoratet. (2015b). *Skoleporten*. Retrieved from https://skoleporten.udir.no/rapportvisning/grunnskole/laeringsresultater/eksamenskarakterer/nasjonalt?rapportid=2&diagraminstansid=1&enhetsid=00&vurderingsomrade=11&underomrade=21&skoletype=0&oversikttypeid=0&fordeling=2&skoletypemenuid=0&sammenstilling=1&kanviseprikking=0&barevisoffentligedata=True&indikator=381&diagramtype=3

Utdanningsdirektoratet. (2016). *Nasjonale prøver*. Retrieved from http://www.udir.no/Vurdering/Nasjonale-prover/#Hva-er-nasjonale-prover

VanDerHeyden, A. M., & Burns, M. K. (2008). Examination of the utility of various measures of mathematics proficiency. *Assessment for Effective Intervention, 33*(4)215–224. doi: 10.1177/1534508407313482

VanDerHeyden, A. M., & Burns, M. K. (2009). Performance indicators in math: Implications for brief experimental analysis of academic performance. *Journal of Behavioral Education, 18*, 71–91. doi: 10.1007/s10864-009-9081-x

VanDerHeyden, A. M., & Burns, M. K. (2010). *Essentials of Response to Intervention.* Hoboken NJ: John Wiley & Sons Inc.

Wong, M. & Evans, D. (2007). Improving basic multiplication fact recall for primary school students. *Mathematics Education Research Journal, 19*, 89–106. doi:10.1007/bf03217451

Yockey, R. D. (2011). *SPSS demystified. A step-by-step guide to successful data analysis* (2. utg.). Upper Saddle River, NJ: Pearson Education, Inc.

Footnotes

[1]The Lin and Kubina (2005) term “fluency” is actually better described as “frequency”, as fluency in the PT-literature is conceptualized as both speed (frequency) and accuracy, and furthermore; a certain frequency with accuracy, only indicates fluency (Binder, 1996). Aside from referring to the Lin and Kubina (2005) study, the term frequency will be used throughout the current study.

Tables

Table 1.
Mean (M), Standard Deviation (SD), Min. and Max. Comparisons.

Variable	Grade level	N	M	SD	Min.	Max.
Numeral writing	5	66	60.58	15.37	25	92
	6	60	69.58	16.39	28	104
	7	63	75.92	21.64	9	108
	8	243	82.55	22.49	0	150
	9	347	84.59	19.65	6	130
Component frequency (digit)	5	66	17.53	9.62	0	45
	6	60	25.58	11.70	0	50
	7	63	33.57	11.68	11	52
	8	243	33.98	15.91	0	106
	9	347	30.24	15.50	0	89
Component accuracy (percent)	5	66	68.21	45.08	0	100
	6	60	62.93	48.36	0	100
	7	63	66.59	47.47	0	100
	8	243	64.95	47.36	0	100
	9	347	56.79	49.18	0	100
Composite frequency (digit)	5	66	2.03	1.73	0	8
	6	60	3.10	2.83	0	10
	7	63	4.38	3.94	0	20
	8	243	4.59	3.73	0	19
	9	347	4.19	4.01	0	33
Composite accuracy (percent)	5	66	57.97	42.26	0	100
	6	60	47.90	35.45	0	100
	7	63	59.77	35.30	0	100
	8	243	62.42	36.39	0	100
	9	347	61.84	37.45	0	100

Table 2.
Comparison with Performance Aim

Variables	Grade level	Aim	Frequency [a]	Percent [b]	M	SD	Min.	Max.
Numeral writing	5	< 40	8	12	33.00	4.96	25	38
		≥ 40	58	88	64.38	12.04	40	92
	6	< 40	2	3	29.00	1.41	28	30
		≥ 40	58	97	70.99	14.77	40	104
	7	< 40	4	6	14.00	7.57	9	25
		≥ 40	59	94	80.12	14.68	42	108
	8	< 40	12	5	17.83	14.99	0	37
		≥ 40	231	95	85.91	17.08	45	150
	9	< 40	9	3	29.11	12.77	6	39
		≥ 40	338	97	86.07	17.55	47	130
Component frequency	5	< 80	66	100	17.53	9.62	0	45
		≥ 80	0	0	–	–	–	–
	6	< 80	60	100	25.58	11.70	0	50
		≥ 80	0	0	–	–	–	–
	7	< 80	63	100	33.57	11.68	11	52
		≥ 80	0	0	–	–	–	–
	8	< 80	241	99	33.48	14.95	0	79
		≥ 80	2	1	94.50	16.26	83	106
	9	< 80	346	100	30.07	15.25	0	73
		≥ 80	1	0	89.00	–	89	89
Component accuracy	5	< 100	25	38	16.08	30.75	0	92
		= 100	41	62	–	–	–	–
	6	< 100	24	40	7.33	24.96	0	96
		= 100	36	60	–	–	–	–
	7	< 100	22	35	4.32	20.25	0	95
		= 100	41	65	–	–	–	–
	8	< 100	97	40	12.19	31.20	0	96
		= 100	146	60	–	–	–	–
	9	< 100	162	47	7.44	24.53	0	98
		= 100	185	53	–	–	–	–
Composite frequency	5	< 40	66	100	2.03	1.72	0	8
		≥ 40	–	–	–	–	–	–
	6	< 40	60	100	3.10	2.83	0	10
		≥ 40	–	–	–	–	–	–
	7	< 40	63	100	4.38	3.94	0	20
		≥ 40	–	–	–	–	–	–
	8	< 40	243	100	4.59	3.73	0	19
		≥ 40	–	–	–	–	–	–
	9	< 40	347	100	4.19	4.01	0	33
		≥ 40	–	–	–	–	–	–
Composite accuracy	5	< 100	37	56	25.03	26.20	0	80
		= 100	29	43	–	–	–	–
	6	< 100	47	78	33.49	25.15	0	80
		= 100	13	21	–	–	–	–
	7	< 100	42	67	39.65	25.31	0	83.33
		= 100	21	33	–	–	–	–
	8	< 100	161	66	43.28	30.17	0	91.66
		= 100	82	34	–	–	–	–
	9	< 100	223	64	40.62	30.34	0	94.11
		= 100	124	36	–	–	–	–

Note: [a] Frequency refers to number of participants in each category, [b] Percent refers to percent of participants in each category.

Table 3.
Variability among Accuracy and Fluency

	Grade level	Component accuracy	Frequency [a]	Percent [b]	M	SD	Min.	Max.
Component frequency	5	< 100 %	25	38	12.96	8.42	0	37
		= 100 %	41	62	20.32	9.32	4	45
	6	< 100 %	24	40	23.54	10.80	0	41
		= 100 %	36	60	26.94	12.21	2	50
	7	< 100 %	22	35	28.73	8.42	14	44
		= 100 %	41	65	36.17	12.43	11	52
	8	< 100 %	97	40	30.75	16.13	0	83
		= 100 %	146	60	36.13	15.44	3	106
	9	< 100 %	162	47	26.86	14.88	0	65
		= 100 %	185	53	33.19	15.56	2	85
	Grade level	**Composite accurancy**	**Frequency [a]**	**Percent [b]**	**M**	**SD**	**Min.**	**Max.**
Composite frequency	5	< 100 %	37	56	1.54	2.05	0	8
		= 100 %	29	44	2.66	0.90	1	4
	6	< 100 %	47	78	2.49	2.45	0	9
		= 100 %	13	22	5.31	3.12	1	10
	7	< 100 %	42	67	3.48	3.22	0	11
		= 100 %	21	33	6.19	4.65	2	20
	8	< 100 %	161	66	4.06	3.65	0	17
		= 100 %	82	34	5.62	3.68	1	19
	9	< 100 %	223	64	3.65	3.54	0	16
		= 100 %	124	36	5.15	4.60	1	33

Note: [a] Frequency refers to number of participants in each category, [b] Percent refers to percent of participants in each category.

Table 4.
Regression analysis between Component and Composite frequency and accuracy.

Component frequency predicting Composite frequency				
Grade level	**N**	**β**	***p***	***R^2***
5	58	.136	.309	.018
6	58	.382	.003	.146*
7	59	.315	.015	.100*
8	231	.427	.000	.182*
9	338	.447	.000	.200*
Component frequency predicting Composite accuracy				
Grade level	**N**	**β**	***p***	***R^2***
5	58	− .284	.031	.080
6	58	.297	.023	.088
7	59	.268	.040	.072
8	231	.117	.076	.014
9	338	.196	.000	.038
Component accuracy predicting Composite frequency				
Grade level	**N**	**β**	***p***	***R^2***
5	58	.104	.436	.011
6	58	.068	.614	.005
7	59	.069	.603	.005
8	231	.064	.336	.004
9	338	.039	.470	.002
Component accuracy predicting Composite accuracy				
Grade level	**N**	**β**	***p***	***R^2***
5	58	−.070	.602	.005
6	58	.119	.374	.014
7	59	.214	.104	.046
8	231	.098	.137	.010
9	338	.019	.730	.000

Note: * = medium effect size

Table 5.
Repeated measures t-test for Numeral writing, Component frequency, Compunent accuracy, Composite frequency and Composite accuracy.

Variable	Grade level	N	January (T1) Mean	January (T1) (SD)	May (T2) Mean	May (T2) (SD)	*t*[a]	*d*
Numeral writing	5	17	57.88	(14.80)	65.00	(15.13)	2.84*	.69^
	6	21	73.43	(15.24)	80.14	(15.63)	2.45*	.53^
	7	22	83.41	(13.65)	87.55	(18.13	2.20*	.47^
	8	131	81.71	(21.67)	88.70	(20.56)	4.74**	.41^
	9	186	85.39	(17.34)	90.61	(20.20)	4.48**	.34
Component frequency	5	17	11.65	(7.049	19.41	(9.23)	4.57**	1.11^^
	6	21	28.52	(12.26)	40.43	(14.02)	5.76**	1.26^^
	7	22	34.09	(13.61)	39.05	(13.69)	2.47*	.53^
	8	131	34.29	(15.87)	36.80	(14.74)	3.20**	.28
	9	186	29.90	(15.60)	36.00	(15.81)	10.71**	.79^^
Component accuracy	5	17	86.77	(18.54)	96.28	(7.17)	2.18*	.53^
	6	21	93.05	(12.63)	97.93	(3.39)	1.65	–
	7	22	98.79	(2.44)	98.58	(3.41)	.21	–
	8	131	95.52	(11.07)	96.11	(9.64)	.51	–
	9	186	93.73	(15.29)	97.74	(4.05)	3.55**	.26
Composite frequency	5	17	1.94	(1.60)	2.18	(1.74)	.51	–
	6	21	3.52	(3.14)	3.81	(3.20)	.33	–
	7	22	5.05	(3.48)	5.55	(3.20)	.56	–
	8	131	4.83	(4.00)	5.42	(4.53)	1.51	–
	9	186	4.31	(4.30)	5.23	(4.65)	2.81**	.21
Composite accuracy	5	17	59.18	(43.40)	61.70	(41.91)	.22	–
	6	21	44.22	(28.97)	50.53	(37.34)	.84	–
	7	22	63.88	(29.70)	66.73	(30.31)	.36	–
	8	131	63.71	(36.28)	62.15	(36.33)	.39	–
	9	186	61.05	(35.86)	66.91	(36.52)	1.77	–

Note: [a]t-test statistics: the degrees of freedom (df) is always N - 1. ES (*d*) are calculated for significant differences only.
* = p < .05, ** = p < .01
^ = medium ES, ^^= large ES

Appendix

Tool Skill Test (Write down 0-9 numerals)

Please use the back of this page if you need extra space.

Component Skill (1 digit x 1 digit multiplication)

8	8	1	2	4	3	8	5	6	1	6	4
x 5	x 4	x 1	x 7	x 8	x 7	x 8	x 6	x 8	x 1	x 0	x 5
2	9	4	0	8	8	2	7	5	9	6	9
x 2	x 5	x 3	x 0	x 6	x 1	x 8	x 1	x 6	x 2	x 2	x 9
1	1	8	1	6	2	8	8	9	7	8	4
x 5	x 6	x 8	x 6	x 9	x 3	x 0	x 3	x 6	x 4	x 3	x 4
4	2	7	1	9	4	1	9	0	6	7	6
x 5	x 4	x 7	x 2	x 5	x 4	x 8	x 9	x 2	x 8	x 4	x 1
0	8	6	1	2	3	8	8	1	9	6	7
x 6	x 2	x 8	x 8	x 1	x 3	x 9	x 3	x 4	x 8	x 3	x 6
8	8	7	4	7	7	6	2	3	3	8	1
x 0	x 4	x 5	x 6	x 8	x 7	x 7	x 1	x 4	x 5	x 4	x 5
1	8	5	2	8	8	4	2	4	4	2	7
x 1	x 8	x 7	x 1	x 8	x 6	x 2	x 4	x 2	x 6	x 4	x 6

2

0 x 4	3 x 8	6 x 6	9 x 2	0 x 6	4 x 2	7 x 2	4 x 1	4 x 6	7 x 3	2 x 8	7 x 8
6 x 7	7 x 0	7 x 9	4 x 3	9 x 0	8 x 2	7 x 8	5 x 8	4 x 5	9 x 4	7 x 4	6 x 8
1 x 7	5 x 9	8 x 3	2 x 4	6 x 5	9 x 8	1 x 4	9 x 2	8 x 6	4 x 5	0 x 4	0 x 2
2 x 5	5 x 3	2 x 5	7 x 3	4 x 8	2 x 7	8 x 3	5 x 9	1 x 1	6 x 2	4 x 6	7 x 1
2 x 3	1 x 4	6 x 8	6 x 7	5 x 8	3 x 8	1 x 5	7 x 8	5 x 4	8 x 3	0 x 1	1 x 2
9 x 2	2 x 1	7 x 1	4 x 7	8 x 4	6 x 0	4 x 1	8 x 5	7 x 8	1 x 0	3 x 6	4 x 1

3

Composite Skill Test (2 or 3 digits x 1 or 2 digits Multiplication)

986 x 83	171 x 59	503 x 9	619 x 71	184 x 88	166 x 48	86 x 9
51 x 68	342 x 76	218 x 96	893 x 17	693 x 75	948 x 14	140 x 59
90 x 77	353 x 49	373 x 27	145 x 66	964 x 81	111 x 63	642 x 55
474 x 46	981 x 1	295 x 20	127 x 41	921 x 79	567 x 47	973 x 55
269 x 31	261 x 52	409 x 94	69 x 4	748 x 49	274 x 6	290 x 92

808 x 51	57 x 88	977 x 1	821 x 7	773 x 67	163 x 39	181 x 8
515 x 37	430 x 22	930 x 61	453 x 32	625 x 82	268 x 89	819 x 56
953 x 64	906 x 76	729 x 27	466 x 21	312 x 74	991 x 70	268 x 96
605 x 22	545 x 42	411 x 8	958 x 8	534 x 67	280 x 97	42 x 95

HØGSKOLEN I OSLO OG AKERSHUS

Leveres sammen med oppgaven i eksamensinnleveringsrommet i Fronter

KLAUSULERING AV OPPGAVE AV ANDRE GRUNNER ENN LOVBESTEMT TAUSHETSPLIKT

AVTALE INNGÅTT MELLOM HØGSKOLEN I OSLO OG AKERSHUS, FAKULTET FOR HELSEFAG OG

STUDENT CATHRINE BERG-MORTENSEN **FØDT** 09.01.86

Studentnummer: 292401 **Studieprogram:** LÆRING I KOMPLEKSE SYSTEMER

Der master- eller bacheloroppgaver ikke inneholder opplysninger undergitt lovbestemt taushetsplikt, kan Høgskolen i Oslo og Akershus med hjemmel i åndsverksloven § 27, jf. offentlighetsloven § 2 og § 6 nr. 6, gi forskere og studenter *innsyn* i oppgaven, men uten at disse kan sitere fra oppgaven.

Studenten kan reservere seg mot at oppgaven gjøres tilgjengelig i høgskolens åpne vitenarkiv, ODA. Dersom denne adgangen til klausulering benyttes, må varighet og begrunnelse angis fra studentens side. Oppgaven vil da bli oppbevart i et lukket arkiv i den oppgitte perioden.

Jeg reserverer meg herved mot at oppgaven gjøres tilgjengelig i høgskolens åpne vitenarkiv, ODA før embargotidens utløp (1-5 år).

Oppgavens tittel:

Main title: On the Relationship Between Component and Composite Multiplication Skills and Natural Frequency Growth

Adresse:	Bråtenvegen 25 a, 1929 Auli
Telefon:	97795080
E-post:	cathrine.bergmortensen@gmail.com

Faglig veileder:	Børge Strømgren	
Institutt:	Atferdsvitenskap	
Jeg tilrår at oppgaven klausuleres	Sted: Kjeller	Dato: 12/5-16
	Veileders underskrift: Børge Strømgren	

Maksimumstid for denne type klausulering er 5 – fem år (embargotid).

Klausuleringen gjelder for 5 år og settes på grunn av:

planlagte innsendinger til fagfellevurderte tidsskrifter

Cathrine Berg-Mortensen
studentens underskrift