EUROPEAN JOURNAL OF BEHAVIOR ANALYSIS 2014, **15**, 225–240 NUMBER 2 (WINTER 2014)

The Use of Precision Teaching to Teach Basic Math Facts

Børge Strømgren, Cathrine Berg-Mortensen, and Linn Tangen
Oslo and Akershus University College of Applied Sciences

A Precision Teaching program for basic math facts was designed for students in the 5th, 6th, and 7th grade level in a regular primary school. Forty-eight typically developing students who were nominated by their teacher as falling behind their classmates in multiplication and division participated. Following a randomization there were two groups of 24 each: one that received Precision Teaching and one that received treatment as usual. The Precision Teaching intervention was put in place for 20 minutes each school day of the week for eight weeks. Tests were performed before and after the program, and also at follow up which was one month after termination of the Precision Teaching intervention. The group receiving Precision Teaching improved more on the tests than did the group receiving "treatment as usual", i.e. the regular mathematic instruction. Furthermore, on an individual level, more students in the Precision Teaching group showed reliable improvement. This study adds to the knowledge of the efficacy and importance of Precision Teaching and mathematic performance monitoring.

Key words: Precision Teaching, mathematics, multiplication, division, math facts, mathematic performance monitoring

---

Mathematics is considered one of the core subjects in the Norwegian compulsory education system, along with reading, writing and computer skills. Mathematics have been prioritized since 2006 with the national initiative called Kunnskapsløftet [The Knowledge Promotion, our translation] (Kunnskapsdepartementet, 2013). Despite this priority, skills in mathematics have not shown much improvement since the implementation of The Knowledge Promotion. Norwegian students have participated nationwide in international assessments of basic skill knowledge, more specifically the Programme for International Student Assessment (PISA) ("Programme for International Student Assessment ", 2013) and the Trends in International Mathematics and Science Study (TIMSS) ("Trends in International Mathematics and Science Study," 2013). Both assessment programs have assessed mathematic proficiency repeatedly over the last decade. There have been four assessments from the PISA conducted in the years of 2003, 2006, 2009 and 2012, and four assessments from the TIMSS conducted in the years of 1995, 2003, 2007 and 2011.

Results from the Norwegian PISA assessment show that the proportion of students performing at the lowest proficiency level has not changed – or improved – from 2003 to 2012 (Kjærnsli & Olsen, 2013). The PISA mathematics proficiency level 1 is stated as:

> At level 1 students can answer questions involving familiar contexts where all relevant information is present and the questions are clearly defined. They are able to identify information and to carry out routine procedures according to direct instructions in explicit situations. They can perform actions that are obvious and follow immediately from the given stimuli (Kjærnsli & Olsen, 2013, p. 54).

Correspondence concerning this article should be addressed to Børge Strømgren, Oslo and Akershus University College of Applied Sciences. e-mail: Borge.Stromgren@hioa.no

Also, results from TIMSS showed only a slight improvement in mathematical skills from 1995 to 2011. In fact, the 2011 TIMSS assessment showed that 37 % of students in the 4th grade scored at or below the Low International Benchmark competence level, while the corresponding proportion was 49 % for students in the 8th grade (Grønmo et al., 2012).

According to Mullis, Martin, Foy & Arora (2012) the TIMSS-definition of Low International Benchmark competence in 4th grade arithmetic is "Students have some basic mathematical knowledge. Students can add and subtract whole numbers. …" (p. 87). In 8th grade mathematics the definition of Low International Benchmark competence is "Students have some knowledge of whole numbers and decimals, operations, and basic graphs. …" (p. 113). The Intermediate International Benchmark competence in arithmetic is defined (Mullis et al., 2012) for the 4th grade as "Students can apply basic mathematical knowledge in straightforward situations. Students at this level demonstrate an understanding of whole numbers and some understanding of fractions. …" (p. 87), and for the 8th grade as "Students can apply basic mathematical knowledge in a variety of situations. Students can solve problems involving decimals, fractions, proportions, and percentages. They understand simple algebraic relationships. …" (p. 113).

The definition of competence levels provided by (Mullis et al., 2012) seem to be in accordance with the competency level aims stated by Utdanningsdirektoratet [The Norwegian Directorate for Education and Training], as their mathematics education plan declare that students by the end of the 4th grade should be able to " … develop and use varied methods of multiplication and division, use these in practical situations and use the standard multiplication table counting in one's head and for solving equations… " (Utdanningsdirektoratet, 2014a), and by the end of the 7th grade should be able to " … find common denominators and carry out addition, subtraction and multiplication of fractions … " (Utdanningsdirektoratet, 2014b).

Thus, the latest PISA and TIMSS findings reflect that many Norwegian students do not meet the competency aims after seven years of schooling. This may hamper competency development in further education (Nasjonalt organ for kvalitet i utdanningen [NOKUT], 2008) and thus have a long term negative societal impact as mathematic proficiency may affect national growth (Hanushek & Kimko, 2000). The Norwegian minister of education stated that "We have a problem with scientific subjects, and it is grave" (Grande & Gjerde, 2013). Furthermore, in the developed and modern western society, a lack of mathematic proficiency may affect individuals' ability to lead a successful life, ranging from reduced levels of income (Murnane, Willett, & Levy, 1995) to making informed decisions about health issues (Reyna & Brainerd, 2007).

During the debate following the PISA and TIMMS findings, suggestions aimed at improving the mathematical proficiency in primary and secondary schools have mainly revolved around improving teacher proficiency in mathematics, which is proposed to be accomplished by the introduction of a new 5-year teacher education program for new teachers as well as further training for teachers currently employed. According to Hattie (2009) this is generally in the low effect range and thus below the zone of desired effects, i.e. below an educational effect size of .40 (Cohen's d). The debate has not resulted in any salient suggestions to focus upon learning strategies in order to improve the students' basic mathematical skills and ways to make sure that more students advance from the lowest proficiency levels to more advanced levels. According to Hattie (2009), teaching strategies are generally of medium effectiveness and thus within the zone of desired effects, i.e. an educational effect size of .40 or more (Cohen's d).

On a general level, the focus is upon developing students' conceptual understanding of mathematical problems in favor of training basic facts before working towards understanding (Kjærnsli & Olsen, 2013). Wu (1999) contended that the dichotomy between basic skills and conceptual understanding is bogus, and that basic skills are a necessary part of understanding; " … The truth is that in mathematics, skills and understanding are completely intertwined. In most cases, the precision and fluency in the execution of the skills are the requisite vehicles to convey the conceptual understanding" (Wu, 1999, p. 1). Also, Gersten et al. (2009) recommend teaching practices such as building fluent repertoires of math fact retrieval along with explicit and systematic instruction including guided practice, corrective feedback and frequent cumulative review.

Precision Teaching (PT) is generally an efficient strategy for teaching primary school students basic and more complex academic skills as well as to remedy academic skill deficits in secondary school (Johnson & Street, 2013; Kubina & Morrison, 2000; Kubina & Yurich, 2012). The PT teaching strategy is general and therefore not dependent upon the subject taught, the learning objectives, or the curriculum. Basic principles are that the learner always knows best (i.e. if the learner does not learn, the learning problem is due to the teacher's lack of appropriate teaching strategies) and that the focus is on observable behavior. Furthermore, PT is not a method of teaching per se, but a teaching strategy comprising five process steps; (1) defining learning objectives (also termed pinpointing), (2) arranging learning materials and procedures to ensure practice, (3) timed practice with frequency counting of performance, (4) daily charting of performance, and (5) teacher review of performance and intervention as needed (Johnson & Street, 2013; Kubina & Yurich, 2012).

PT has been used to diagnose, or pinpoint, mathematic proficiency, and to build mathematic proficiency both in primary and secondary schools. In PT, true proficiency is termed fluency, which is determined by both accuracy and speed. So being accurate is not enough; one also needs to be accurate with sufficient speed. Children who can solve a number of arithmetic problems with 100 % accuracy will not be considered fluent if it takes the child too long to solve them (Kubina & Morrison, 2000).

In order to generate fluency, certain frequency aims, or performance standards, have been proposed. Johnson and Street (2013) propose frequency aims of 160-130 correct digits per minute when writing numbers 0 - 9, and 100-80 correct answers per minute when seeing single digit problems (e.g., 3 × 9) and writing answers, which they term Math Facts. Math problems including multiple digits (e.g., 13 × 9, or 27 × 17) they term Computation, and they aim for 60-65 answered digits per minute. Haughton (1972) proposed a frequency aim of 40-50 correct problems (and 80 correct digits) as appropriate for basic math fact computation, as lower frequencies would slow the learner down as the student progressed through the curriculum or the amount of time spent working increased (in PT the ability to keep up performance is termed Endurance). Also, Binder (1996) suggested that frequencies of between 50-70 are necessary in order to maintain performance during short (two minutes) and long (four minutes or more) intervals. The PT focus upon universal screening and pinpointing also fits nicely within the Response to Intervention framework, where it is recommended that these evaluations be done three times each school year in order to identify learneras at risk for learning problems (Lembke, Hampton, & Beyers, 2012).

VanDerHeyden and Burns (2009) performed a brief experimental analysis with 2nd to 5th grade students in which they screened math proficiency in multiplication, division (Math Facts) and Fact Families. They found that 4th and 5th graders showed an average frequency of 42.79 correct division problems

(÷ 0 to ÷ 12) and 46.95 correct multiplication problems (× 0 to × 12). Also, the average frequency of correct answers for Fact Families (mixed ÷/× 0 to ÷/× 12) was 43.15. They concluded that frequency levels that predicted retention were 61 (÷ 0 to ÷ 12), 63 (× 0 to × 12), and 64 (÷/× 0 to ÷/× 12).

Lin and Kubina (2005) performed a screening of fluency levels in 5th graders in order to look at the association between fluency and accuracy in multiplication. They aimed for 80-120 correct digits per minute with single digit problems (Math Facts), and 40-60 correct digits per minute for multiple digit problems (Computation). Results showed a mean frequency of 57.01 correct digits per minute for single digit multiplication problems (Math Facts) – only 14 % of the participants performed at the aim of 80-120. For the multiple digit problems (Computation), results showed a mean frequency of 18.71 correct digits per minute and only 3.2 % of the participants performed at the aim of 40-60. They concluded that accuracy was not sufficient to master new and more complex calculation skills, or Computation, and that fluency have a significant role in this respect.

We have found two studies demonstrating the use of PT in order to build proficiency in basic math facts with typically developing children in a regular school setting. Chiesa and Robertson (2000) delivered a 12 week PT-program to 5 primary school children who – by teacher nomination – were identified as lagging behind in basic multiplication and division curriculum progress. The remaining 25 students in the class received treatment as usual, that is regular mathematic instruction. Both a pre-intervention and a post-intervention test were employed for the 5 students receiving the intervention, and also for the remaining 25 students in the class. The test consisted of Computation of division problems, specifically two-digit numbers divided by one up to five (e.g., 70 ÷ 1 to 75 ÷ 5), which were performed during one minute probes.

On the pretest, all 5 students receiving the intervention performed similar to or worse than the 25 remaining students in the class. The mean score for the intervention group was 1 correct response per minute while the mean score for the remaining students was 3.7 correct responses per minute. Following the intervention, all 5 students receiving the intervention out-performed 24 of the 25 remaining students in the class receiving treatment as usual with the mean score for the intervention group being 13.2 correct responses per minute, while the mean score for the rest of the class was 4.2.

Gallagher (2006) also carried out a 12-week PT program in order to teach multiplication timetables to primary school students. The program was used in a classroom where 8 students – nominated by class teacher as lagging behind the rest - followed the PT-course format (PT-group) and where the remaining 15 students in the class followed a Transfer Test[1] preparation course with regular instruction format (TT-group). A pre-intervention and a post-intervention test were employed for both groups. The test comprised 80 randomly selected multiplication problems (Math Facts) from × 1 to × 6 (e.g., 5 × 4, 8 × 6). At the pre-intervention, test the PT-group mean score was far below the TT-group mean score, with a mean of 16.38 correct answers for the PT group and a mean of 26.86 for the TT group. At the post-intervention test, the PT-group mean score had improved, a mean of 22.75, whereas the TT-group mean score now was 27.13. At both pre- and post-intervention the TT-group showed a higher mean score than the PT-group did, but the point here is that the regular instruction format (TT group) did not show any improvement, while the students in the PT-instruction format did.

The current study investigated the effects of a Precision Teaching intervention in math

Fotnote 1. In UK the Transfer Test examination tests a student's ability to solve problems using verbal reasoning and mathematics, and it was used to determine which type of school the student should be placed in after primary school; grammar, secondary or technical school.

facts (multiplication and division) when delivered in addition to a treatment as usual procedure (regular classroom-based mathematics teaching) as opposed to a treatment as usual procedure alone. We were looking to test four hypotheses. First, we predicted that more students receiving Precision Teaching would double their test score from a pre-intervention test to a post-intervention test and follow-up test than would students receiving treatment as usual. Second, we predicted that more students receiving Precision Teaching would show statistically and clinically reliable change at post-intervention and follow up tests than would students receiving treatment as usual. Third, for the Precision Teaching participants we predicted that a frequency aim of 50 correct answers per minute or more would predict Endurance and Stability in the same frequency range, and that a lower frequency aim would not. Fourth, we predicted that the Precision Teaching participants would rate their participation in the PT-intervention favorably.

Method

Participants

Typically developing students from the 5th, 6th, and 7th grades in a regular primary school participated (n = 48). All participants had been introduced to basic multiplication and division math facts via regular classroom teaching procedures. Class teachers nominated students who they regarded were lagging behind the rest of the class in basic multiplication and division skills. Written and informed consent was obtained from nominated students and their parents/guardians.

Nominated students were randomly allocated to either a PT-intervention (PT) or treatment as usual (TAU) group. In order to force equal sample sizes we used a restricted random assignment method (Shadish, Cook, & Campbell, 2002). Following the random allocation, each group consisted of 24 participants. The TAU group comprised 24 students: six girls and three boys from the 5th grade, six girls and three boys from the 6th grade, and four girls and two boys from the 7th grade. The PT group comprised 24 participants: four girls and five boys from the 5th grade, four girls and five boys from the 6th grade, and four girls and two boys from the 7th grade.

Setting

All tests and PT training sessions were conducted at the participants' regular primary school in a designated training room. The room was usually assigned to teaching digital skills and had 10 desks and a cupboard. Seven of the desks had computers; they were not used and were moved towards the wall side of the desk with the screen facing the wall.

PRE, POST and FU Test for Both Groups

Stimuli and materials. Following randomization, we constructed a test in order to obtain Pre-intervention (PRE), Post-intervention (POST) and Follow-Up (FU) measures of correct answers per minute (CAPM) in both multiplication and division. The test was based upon a typical test that had previously been used at the participating school. It consisted of 5 worksheets with 250 mixed multiplication and division problems (50 mixed problems per work sheet) printed on white A4 sheets. The mixed problems ranged from multiply by one (M1) and divide by one (D1) to multiply by 12 (M12) and divide by 12 (D12). Test sheets with corresponding answer keys were generated and downloaded from the website http://themathworksheetsite.com/.

Procedure. Participants from both groups completed the test before the PT-instruction commenced (PRE), after PT-instruction was terminated (POST), and one month after the termination of the PT-instruction (FU). Participants were given a maximum of 20 minutes to complete the test. Some participants finished before the maximum time limit, and for those participants, the PT-instructors noted the amount of minutes spent completing the test.

During the course of the test, there were up to 10 participants and four PT-instructors present at anyone time.

Dependent measures and reliability. The participants' written answers (i.e., corrects and learning opportunities) were independently scored and recorded by two PT-instructors. Each participant's number of correct answers was divided by the amount of time that participant sat for the test in order to calculate correct answers per minute (CAPM). All reported scores are expressed as correct answers per minute (CAPM). Inter Observer Agreement (IOA) was calculated for all scores by means of the following formula: number agreed/(number agreed - number disagreed). The IOA scores were 99.8 % (SD 0.45, range 98 - 100) for PRE, 99.9 % (SD 0.33, range 98.5 - 100) for POST, and 99.9 % (SD 0.33, range 98.5 - 100) for FU.

Design and data analysis. For the purpose of analyzing initial group differences we performed a Mann-Whitney U test which showed that both groups scored equally at PRE, U = 261, p = .950. The PT group had a mean rank of 21.63 and the TAU of 21.39. For the purpose of analyses of between group differences we used non-parametric chi-square statistics to analyze the amount of participants who had doubled their test score from PRE to POST and from PRE to FU.

In order to analyze variability and change for each participant in both groups, we used the Reliable Change Index (RCI) (Christensen & Mendoza, 1986; Jacobson & Follette, 1984; Jacobson & Truax, 1991; Speer, 1991). The RCI makes it possible to determine which test scores showed a clinically and statistically significant improvement or decrement from PRE to POST and from PRE to FU. The initial standard deviation for both groups scores combined, together with the TAU group PRE to POST and PRE to FU test-retest reliabilities were used for calculating RCI by the formula provided by Jacobson and Truax (1991).

From the TAU group, one participant dropped out from POST; consequently, chi-square analyses of group test scores and RCI analyses of individual test score changes for this group was calculated with 23 participants. From the PT group, two participants withdrew from the PT-procedure and three participants dropped out from POST or FU. Therefore the chi-square analysis of PT group test scores was calculated with 19 participants, and the RCI analysis of individual score changes with 22 participants.

Precision Teaching for the PT Group Participants

Stimuli and materials. Each PT group participant had his or her own folder containing Timings Charts (TC) and Standard Celeration Charts (SCC) (one SCC for multiplication and one SCC for division), and a log form to be filled in at the end of each session. We used separate work sheets for each multiplication and division table, i.e., from multiply by 1 (M1) to multiply by 12 (M12) and from divide by 1 (D1) to divide by 12 (D12). Work sheets with corresponding answer keys were generated and downloaded from the website http://themathworksheetsite.com/. Each worksheet consisted of 50 problems, and was printed on white A4 size paper. In addition to work sheets, each participant had a timer, a pencil and the answer key. During Stability tests, which are timed practices with distractions (e.g. music played), participants used headphones with music played from a mobile phone.

Procedure. The present study was conducted over a period of eight weeks during the school spring semester, from late January to mid March with a one week pause in mid February due to school winter holiday. Training sessions were conducted each school day during regular school hours when students' regular schedule permitted. Participants were thus pulled out during different subjects, usually Norwegian, English, Social Studies, Science, RLE (religion, philosophy of life and ethics) and Mathematics.

During the course of the study, there were from one to 10 participants and from two (the second and third author) to four PT-instructors present at anyone time.

Precision Teaching instruction. Previous to PT-instruction in multiplication and division, we conducted an introductory lesson where participants were informed about and received instruction in the timing procedures and use of materials such as the TC and the SCC. The learning channel was see problem / write answer throughout the study, meaning that the students saw the math problem on the work sheet and were required to write their answer on the work sheet.

M1/D1 instruction. All PT group participants started with M1 and D1 problems, and kept at that level until they reached their Personal Best (PB) frequency limit or a initial preset overall frequency aim of 70 CAPM based upon the suggestions from Haughton (1972) and Binder (1996). The M1/D1 functioned as a prolonged introduction as it made it possible for participants to get familiar with the timings procedure and materials in addition to scoring their SCC. Also it provided an opportunity for participants to experience improvements in rates, or celeration, from timing to timing. Finally it provided an opportunity for the PT-instructors to assess each participant's Personal Best (PB) frequency level potential, when their celeration "flattened out". Each participant's PB score at M1/D1 served as a guide for future frequency aims and intervention procedure decisions for that participant. Participant's M1/D1 Personal Best, Stability and Endurance scores are depicted in the M1 CAPM and D1 CAPM columns in Table 1.

Table 1

PT Group Participant's ID, Grade Level, Total Number of Sessions, M Training Details, M1 Scores, D Training Details, and D1 Scores.

			M training			M1 scores			D training			D1 scores		
ID	Grade	Sessions	Timings	Level	Test	PB	Stability	Endurance	Timings	Level	Test	PB	Stability	Endurance
PT01	5	25	87	5	M3	60	82	66 (1)	94	5		64	72	67
PT02	5	27	116	4	M2	46	60 (2)	39	137	4		63	60	45
PT03	5	24	93	4	M2	55	52	45	122	4		61	62	58
PT04	5	27	150	6	M3	50	62	55	116	5		45	58	52
PT05	5	27	116	6	M3	58	80	65	119	5	D3	65	64	56
PT06	5	25	106	4	M2	58	60	51	110	4		61	56	49
PT07	5	26	142	5	M3	63	62	61	108	4		81	76	63
PT08	6	26	105	6	M4	63	72	67	145	5		60	74	69
PT09	6	19	42	4	M2	61	52 (2)	59	117	3	D2	65	60	62
PT10	6	25	116	6		55	86	62	94	5	D3	66	72	67
PT11	6	28	116	6		65	76 (4)	59 (1)	103	5	D4	30	72	60
PT12	6	27	91	7	M7		52	39	154	8	D4	61	72	54
PT13	6	28	95	10	M2				148	10	D2			
PT14	6	31	108	8	M4	60	98 (2)	88	124	9	D7	73	100	75
PT15	6	27	101	8	M3	59	72 (4)	64 (1)	86	7	D7	65	74	61 (1)
PT16	6	26	88	6	M4	63	76	70 (1)	137	6		63	84	75
PT17	7	20	62	3	M3	56	82	71 (1)	126	4		41	84	71
PT18	7	19	90	7	M4				96	7	D7			
PT19	7	26	111	7	M3	67	84	66	125	6		61	70	60
PT20	7	25	133	6		44	74 (2)	66 (1)	93	5	D4	67	82	57
PT21	7	24	162	4	M3	57	76	57	79	3		63	64	55
PT22	7	30	106	5		60	48	49 (1)	160	6	D3	45	82	50

Note: Sessions are total number of sessions. Timings are total number of timings in M and D. Level is M/D level reached during training. Test is M or D level at which later Personal Best (PB) Stability (S), and Endurance (E) tests were performed. Numbers in parentheses represent learning opportunities.

Session preparation and closure. PT-instructors prepared each session by means of consulting logs from the previous training session, placing the participants' training materials on their designated desk and collecting participants from their regular classroom session. After session closure, the PT-instructors followed participants back to their class. PT sessions lasted 25 minutes.

Session procedure. All practice was timed, usually with 30 second timings, with a maximum of 10 timings per participant per session. Each participant set and started his or her timer and answered as many math problems as possible during the timing. Following the timer buzz, participants stopped writing answers, checked their answers with the answer key, counted correct responses and learning opportunities, and filled in on their TC. Any math problem that had been skipped during timing counted as a learning opportunity. At the end of the session, each participant plotted his or her best session score on the SCC. PT-instructors were available if help or new material was needed. After the session, all work sheets, TCs and SCCs with pupil recorded corrects and learning opportunities were reviewed by the PT-instructors. The M training columns in Table 1 depicts the participant's number of M timings, M level reached, and M level at which later tests of Stability and Endurance were performed. The D training columns in Table 1 depicts the participant's number of D timings, D level reached, and D level at which later tests of Stability and Endurance were performed.

Instruction from M2/D2 and throughout. Participants advanced to the next level (i.e., from M2 to M3, or from D2 to D3) when they reached their initial M1/D1 score, or when their scores "flattened out" in three or more consecutive timings during a session, indicating a PB for that level.

Intervention procedures. Intervention procedures were conducted if celeration towards the initial PB M1/D1 scores was not met, or if a celeration from timing to timing was lacking. One of three general principles were employed: (1) a shortened timings interval, (2) simplify the task with fewer math problems or learning opportunity tasks only, or (3) advancement to the next level despite lower frequency performance than initial PB if (1) and (2) proved unsuccessful. Some participants received extra un-timed practice with a PT-instructor using the learning channel see problem / say answer, and some participants earned awards (e.g., listening to music) following an agreed upon number of timings.

Stability and endurance testing. We performed Stability and Endurance tests at different times for different students by using the same work sheets as in ordinary timings, see Table 1, Test columns for details. Endurance tests were timed to 90 seconds, and Stability tests were timed to 30 seconds with music in earphones as distractor.

Dependent measures and reliability. All reported scores are correct answers per minute (CAPM). We initially aimed for an overall frequency of 70. However, some PT-group participants failed to reach that frequency aim and/or celeration, and if their celeration flattened out, they could still advance to the next level as explained in the Intervention procedures section. We still wanted to see frequencies of at least 40-50.

Personal best, stability and endurance. Twenty-two participants completed the PT intervention, and 21 received Stability and Endurance tests at different M or D levels throughout the intervention (See Table 1). For these 21 participants, IOA was calculated for actual Personal Best, Stability and Endurance scores by means of the following formula: number agreed/(number agreed - number disagreed). A total of 81 IOA scores were calculated, all showed 100% IOA.

Post PT-intervention questionnaire. Participants were asked to fill in a questionnaire at time of PT-instruction termination in order to judge their perception of the PT-instruction.

18 participants answered the questionnaire. Among the questions asked were: (i) "Following PT-instruction, do you think it has been useful?", (ii) "Would you consider participating in PT instruction for another math skill (e.g., fractions)?", (iii) "Would you consider recommending PT-instruction to someone else?", and (iv) "Overall, how satisfied are you with the PT-instruction?"

Results

PRE, POST and FU Test for Both Groups

As described in the Method section, the groups' mean CAPM test scores did not differ at PRE. Results from subsequent tests, POST and FU, show that both groups made progress, and that the PT group made larger progress. The PT group had the following mean CAPM test scores at PRE; POST and FU; 5.16 (SD = 2.87), 8.04 (SD = 4.59) and 9.10 (SD = 6.0). Corresponding mean CAPM test scores for the TAU group were 5.07 (SD = 2.21), 6.21 (SD = 2.05), and 7.17 (SD = 2.86).

As can be seen in Figure 1, in the PT group, 5 out of 19 students doubled their test score from PRE to POST, while 1 out of 23 in the TAU group did, a statistically significant difference, χ^2 (1, 42) = 4.10, p = .043. The effect size is medium, Cramer's V = .31, p = .043. At FU, the difference was maintained with 9 out of 19 students in the PT group doubling their test score from PRE to FU, while 2 out of 23 in the TAU group did, a statistically significant difference, χ^2 (1, 42) = 8.05, p = .005. The effect size is medium, Cramer's V = .44, p = .005.

The overall picture from the chi-square analysis is maintained when each group member's test score changes are considered individually. Changes from PRE to POST are displayed in Figure 2 with the RCI marked as horizontal dotted lines. Specifically, from PRE to POST the RCI is ± 3.63, indicating that test score changes of greater magnitude indicate a reliable change. For the PT group, seven participants showed reliable improvement and one participant showed reliable decrement. For the TAU group one participant showed reliable improvement and none showed reliable decrement.

From PRE to FU the RCI is ± 3.16, indicating that test score changes of greater magnitude indicate a reliable change.

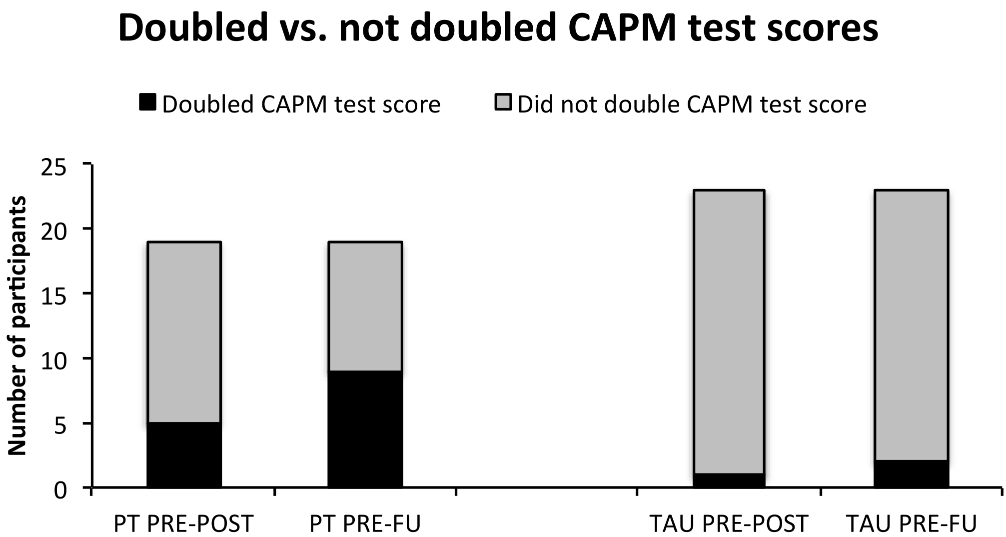

Figure 1. Doubled test scores (x2) from PRE to POST and from PRE to FU for the PT and TAU groups.

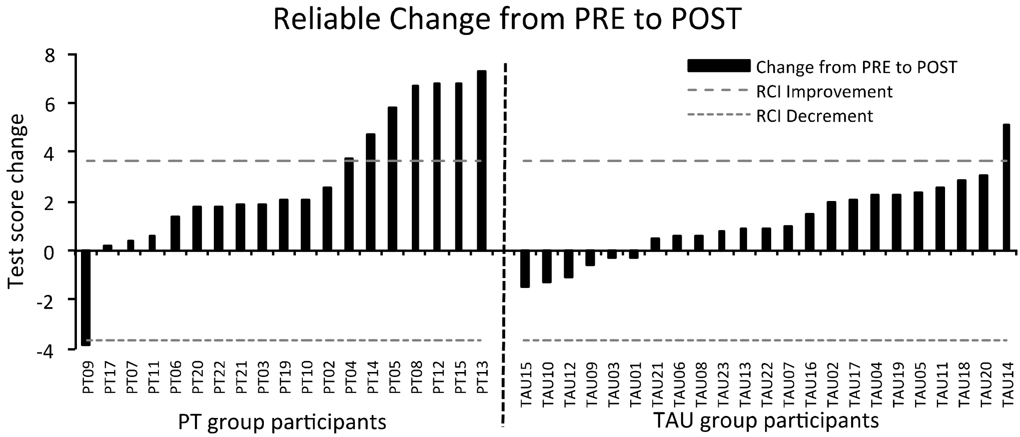

Figure 2. Reliable Change scores from PRE to POST for the PT group TAU group participants.

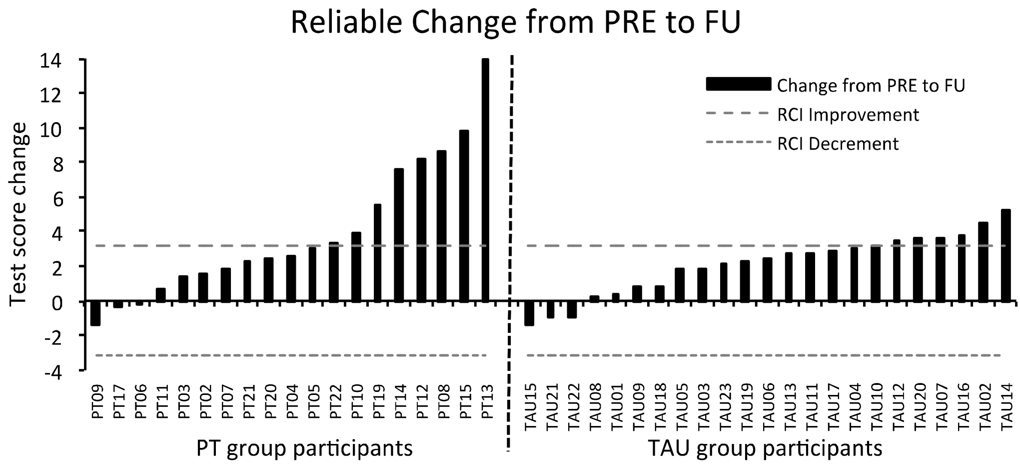

Figure 3. Reliable Change scores from PRE to FU for the PT group TAU group participants.

For the PT group, eight participants showed reliable improvement and none showed reliable decrement. For the TAU group seven participants showed reliable improvement and none showed reliable decrement.

Individual Scores for the PT Group

Multiplication. We obtained Personal Best (PB), Stability (S) and Endurance (E) scores for both M1 and later M tests for 15 participants. Details are depicted in Figure 4 upper left panel (M1) and upper right panel (later M test). At M1 participant PT02 had a PB score of below 50 but scored higher on S and lower E. PT02 maintained the PB score at a later M test but scored far below on S and E. 6 participants (PT04, PT03, PT17, PT21, PT05, PT06, & PT15) had a PB score of between 50 and 60 at M1 and all but for two (PT03 and PT06) improved at S and E. Only two (PT06 and PT15) maintained their PB score at a later M test, and only two (PT04 & PT17) had an S score above their PB score. 7 participants (PT01, PT14, PT09, PT07, PT08, PT16 & PT19) had a PB score between 60 and 70 at M1, and all but for one (PT09) maintained or improved their M1 S and E scores. At a later M test none maintained any of their M1 scores.

Division. We obtained PB, S and E scores for both MD1 and later D tests for 9 participants. Details are depicted in Figure 4 lower left panel (D1) and lower right panel (later D test). At D1 two participants (PT11 & PT22) had a PB score below 50, but higher S and E scores. At a later D test PT11 had a PB score above the D1 PB score, but did not maintain that score at S and E. PT22 decreased from D1 at all later D tests. 7 participants (PT12, PT05, PT09, PT15, PT10, PT 20, & PT14) had a PB score above 60 at D1. All but for one (PT12) maintained or improved at S, and all but for one (PT05) maintained or improved at E. At later D test only three (PT12, PT09 & PT15) maintained or improved their PB score, and only one (PB14) improved at S.

Post PT-intervention Questionnaire

18 PT group participants answered the post-intervention questionnaire. 14 out of 18 rated the PT-instruction as "rather" or "very" useful. 9 out of 18 would have participated for another skill (e.g., fractions). 14 out of 18 would consider recommending this PT-instruction to someone else, and 15 out of 18 were "rather" or "very" satisfied with the PT-instruction?"

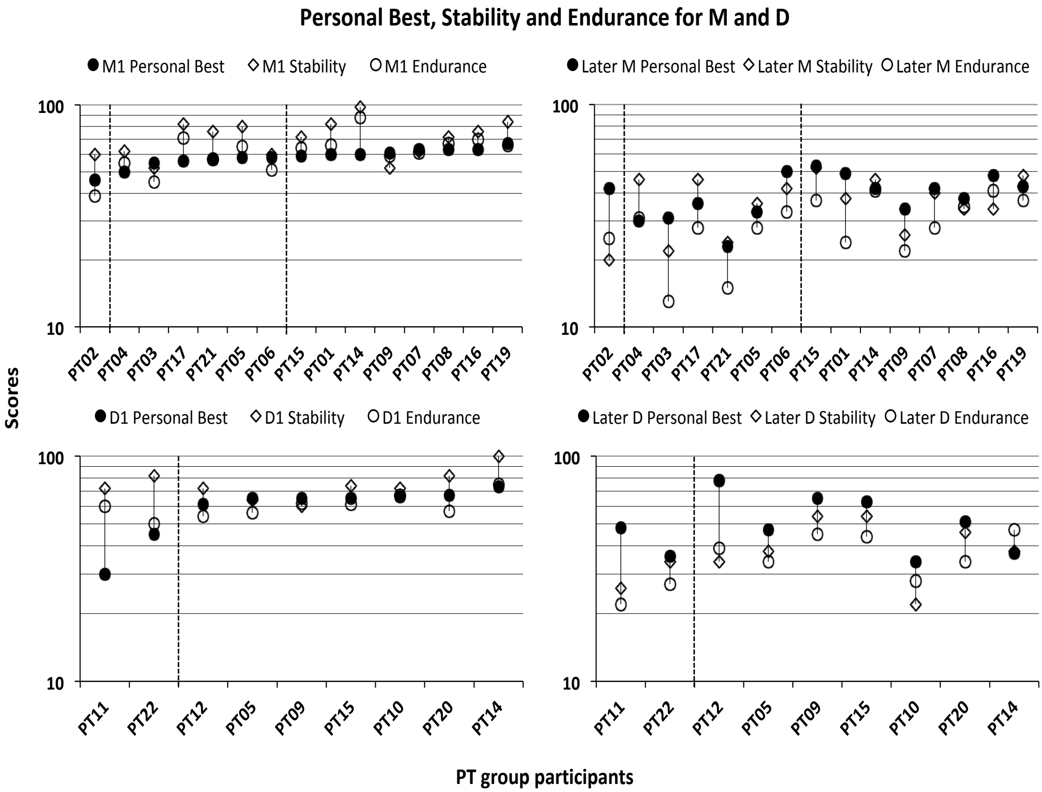

Figure 4. Personal Best, Stability and Endurance score details at M1/D1 and later M/D tests for some PT group participants.

Discussion

The purpose of the current study was to evaluate an eight week PT intervention in order to teach basic math facts. We were looking to test several hypotheses, which will be summarized and commented at the group and individual level.

PRE, POST and FU Test for Both Groups

We predicted that more students receiving PT would double their test score from a pre-intervention test to a post-intervention test and follow-up test than would students receiving TAU. As the difference was statistically significant both from PRE to POST and from PRE to FU, this prediction held true. We also predicted that more students receiving PT would show statistically and clinically reliable change at POST and FU tests than would students receiving TAU. Results indicate that this prediction also held true; more PT group students than TAU group students showed reliable change at POST and FU.

A possible weakness with the present study is that teachers nominated participating students as lagging behind the rest of the class. The pretest data from Chiesa and Robertson (2000) and Gallagher (2006) studies, indicated that teacher nomination is a questionable selection criterion alone as some students in their treatment as usual groups scored similarly to students in the Precision Treatment groups at pre-training tests. Furthermore, they did not use randomization when selecting participants to either the Precision Teaching or the treatment as usual groups, but relied instead on teacher nominations alone. Thus, the present study represents an improvement in the field, as we used a randomized assignment to either PT or TAU among those students that were nominated. We have not found studies involving students in primary school where randomized allocation to a PT-intervention or treatment as usual has been employed.

We think that the randomization procedure rules out some of the weakness represented by the nomination uncertainty. We were not able to perform a screening procedure prior to the teacher nomination; we were only allowed to perform a pre-intervention test after teacher nomination, parent/student informed consent and our randomization. The groups did not differ at the pre-intervention test, but we still would have preferred to do an initial screening of all students in the 5th, 6th, and 7th grade at the participating school. By doing a screening, we could have been more certain that the invitation to participate actually was aimed towards the lower performing students.

The test scores from the present study represent Math Facts; thus they may be compared to the Math Fact screening and test scores found in previous studies (Gallagher, 2006; Lin & Kubina, 2005; VanDerHeyden & Burns, 2009). The mean test scores from PRE, POST and FU are lower than the mean score of 57.01 reported by Lin and Kubina (2005), lower than the mean score of 43.15 reported for 4th and 5th graders by VanDerHeyden and Burns (2009), and also lower than the mean scores of 16.38 to 27.13 reported by Gallagher (2006). This may reflect that we managed to recruit low performers, and that teacher nominations were accurate; but again - since we were not allowed to perform the test for all students, we cannot say this with certainty.

The fact that the test scores are lower in the current study may also be a function of the test procedure. Participants could sit for 20 minutes, which may have caused a decline in speed of responding during the test period. VanDerHeyden and Burns (2009) used two minute probes while Lin and Kubina (2005) used a one minute assessment timing procedure which would be less prone to a decline in speed. The test scores may also be a case of participants being conscious about making any mistakes.

They preferred taking their time in order to be accurate and thus avoid mistakes. This preoccupation with failure avoidance was also salient in later timings with the PT group participants; this may be seen from the M1/D1 scores in Table 1, which show very few occurrences of learning opportunities or skipped problems. Furthermore, the nature of the test material with randomly mixed problems up to ×12 and ÷ 12 may have contributed to the low scores, as participant may have faced math problems they did not know (e.g., 12 ÷ 9), which in turn may have caused some to spend too much time trying to solve the problem in order not to get a mistake, or caused some to give up all together as some did quit before the allowed time given to sit for the test was up.

Thus, test scores in the current study may not be directly comparable to test scores in the above-mentioned studies. Still, we think, the important issue here is the improvement in test scores from PRE to POST and also from PRE to FU. 5 out of 19 students in the PT group doubled their scores in eight weeks (from PRE to POST). Kubina and Yurich (2012) classify a growth of times two as massive if it is achieved in a week. A growth of times two in eight weeks will be around times 1.25 in a week, which is classified as acceptable. Thus, for 5 of the PT group participants the growth from PRE to POST can be viewed as acceptable or better. Data from the reliable change calculation support this notion that a portion of the PT group participants did achieve a noteworthy improvement form their participation in the PT instruction.

Individual Scores for the PT Group

For the PT group students, we assumed that a frequency aim of 50 correct answers per minute or more would predict similar Endurance and Stability scores, and that a lower frequency aim would not. Binder (1996) indicated that number-writing frequencies of below 70 per minute probably would result in performance decrements during long timing intervals (i. e., timings of 4 minutes or longer). For timings up to 2 minutes, initial frequencies of 50 per minute would suffice to maintain performance, and initial frequencies of 40 would show some decrements.

Figure 4 show individual PB scores for some participants on M1 and a later M timing (top panels) and on D1 and a later D timing (bottom panels). The PB scores on M1 and D1 can be regarded as being close to number writing only as there is not much of a multiplication or division task other than repeating the original number multiplied or divided by 1. As such, the M1/D1 PB scores may serve as a baseline for a participant's number writing proficiency. The PB scores are still below the aim proposed for number writing by Johnson and Street (2013), Binder (1996), and Haughton (1972). At both M1 and D1, most participants scored above 50, and maintained their scores at Stability and Endurance tests. For later M and D levels we saw a marked reduction in PB scores and also in Endurance scores. Still, 8 participants scored between 40 and 50 in the later M PB and 6 in the later D PB. As we did not perform Stability and Endurance tests at several time intervals, we cannot say whether this level was sufficient to maintain the same level of PB scores throughout the PT intervention; but from the reliable change differences along with the different M and D levels reached for each participant, we can conclude that some probably did not.

The lack of PRE to POST reliable improvement for 12 out of 19 participants may be due to procedural or organizational reasons. Novice trainers carried out the PT procedure, and they found it difficult to overcome the participant's focus on avoiding errors, or learning opportunities. This can be seen by the fact that Endurance scores were lower than PB scores, which would be expected with longer timings. Bear in mind that skipping a task also counted as a learning opportunity; so with longer timings they took more time to ponder with the math problem or find the answer in the available answer key before they wrote an answer and moved on. Some were also anxious not to look at the answer key during the test as this usually is considered cheating. Stability tests, which hade the same timing as the PB score, did not show the same overall score drop as Endurance, because the condition did not leave them more time to ponder with the problem or looking at the answer key before writing an answer.

The trainers used interventions such as shortened timings and/or simplified tasks or learning opportunity tasks only but still some did not improve in the current or subsequent M or D levels. Extra time spurts with the see problem / say answer were used with some, again with sparse improvement in the current or subsequent M or D levels. It may have been that these interventions should have been used to a larger degree with some participants. The organization of the PT intervention may in part explain why this was not done to a larger extent as there were up to 10 participants and 2-4 PT-instructors in the designated training room at any one time. Participants kept track of their own timings and scores and the PT-instructor reviewed all timings charts and logs after sessions, thus it is possible that some opportunities to intervene were lost or put in place too late.

Post PT-intervention Questionnaire

We also assumed that students participating in the PT group would rate their participation in the PT-intervention favorably, and most did. Only half would consider participating using this same procedure for another math skill. This may be due to the organization of the questionnaire; it was administered in connection with the POST test. These answers do not constitute a measure of social validity, which usually is done before an intervention commences but merely a measure of participants' point of view. In behavior analysis these measurements, of social validity before an intervention commences and participants' judgment after termination, may be overlooked as the main focus often is upon communicating the effectiveness of the intervention. In order to advocate interventions based upon behavior analysis in regular settings, be it regular schools with typically developing students or other regular recipients, it is also important to communicate that the recipients themselves judge the intervention as favorable.

Recommendations

We recommend performing a universal screening in order to identify learners at risk for learning problems (Lembke et al., 2012), possibly by means of a brief experimental analysis (VanDerHeyden & Burns, 2009) or a one minute timed screening (Lin & Kubina, 2005). This will help identifying students at risk and furthermore help to pinpoint each student's level of proficiency, thus ensuring that students' receive individualized intervention at the proper level, as was the case with previous studies (Chiesa & Robertson, 2000; Gallagher, 2006). Also, developing math fact tests with a narrower array of problems, for example, with tests of say from (×1 to ×3), and then from (×4 to ×6) and so forth will allow for a more fine-grained pinpointing.

We also recommend Endurance and Stability testing before every shift (e.g., from M1 to M2, from M2 to M3 and so on), this will help assure that the performance does not drop as the students move to more complex mathematical skills. Furthermore, we recommend the approach described by Johnson and Street (2013) in which they train say ×2, the ×3, then a mix of ×2 and ×3 before moving on to ×4 and so forth, again with Endurance and Stability tests before each shift. This strategy may also be helpful in order for PT-instructors to keep pace with each student's progress and put in place necessary interventions in due time.

Another recommendation is to spend enough time in instructing students in the use of the Timings Chart and the celeration aim line, and teaching students to solicit help if their celeration does not meet their aim line over the course of the timings. This will help the PT-instructor to intervene early enough to maintain the celeration aim. A last recommendation is that PT-trainers review each student's Timing Chart and log together with the student at the end of the session. This can make it easier to plan and put in place proper interventions for each student, and also teach the student to review their own data and make decisions for their own progress, thus contributing to self-directed learning.

References

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst, 19*, 163–197.

Chiesa, M., & Robertson, A. (2000). Precision Teaching and Fluency Training: Making maths easier for pupils and teachers. *Educational Psychology in Practice: Theory, Research and Practice in Educational Psychology, 16*, 297–310. doi: 10.1080/713666088

Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*, 305–308. doi: 10.1016/S0005-7894(86)80060-0

Gallagher, E. (2006). Improving a mathematical key skill using precision teaching. *Irish Educational Studies, 25*, 303–319. doi: 10.1080/03323310600913757

Gersten, R., Beckmann, S., Clarke, B., Foegen, A., Marsh, L., Star, J. R., & Witzel, B. (2009). Assisting students struggling with mathematics: response to intervention (RtI) for elementary and middle schools (NCEE 2009-4060). Washington DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Grande, A., & Gjerde, A. (2013, December 3). Vi har et realfagsproblem, og det er alvorlig. [We have a problem with scientific subjects, and it is grave]. *Dagens Næringsliv*. Retrieved from http://www.dn.no

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, Norway: Akademika.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review, 90*, 1184–1208.

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.

Jacobson, N. S., & Follette, W. C. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336–352. doi: 10.1016/S0005-7894(84)80002-7

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12–19. doi: 10.1037/0022-006X.59.1.12

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, Norway: Universitetsforlaget.

Kubina, R. M., & Morrison, R. S. (2000). Fluency in education. *Behavior and Social Issues, 10*, 83–99. doi: 10.5210%2Fbsi.v10i0.133

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

Kunnskapsdepartementet. (2013). *Kunnskapsløftet* [The Knowledge Promotion]. Retrieved 05.05.2013, 2013, from http://www.regjeringen.no/en/dep/kd/Selected-topics/compulsory-education/Knowledge-Promotion.html?id=1411

Lembke, E. S., Hampton, D., & Beyers, S. J. (2012). Response to intervention in mathematics: Critical elements. *Psychology in the Schools, 49*. doi: 10.1002/pits.21596

Lin, F.-Y., & Kubina, R. M. (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

Mullis, I. V. S., Martin, M. O., Foy, P., & Arora, A. (2012). *TIMSS 2011 International results in mathematics*. Chestnut Hill, MA: MSS & PIRLS International Study Center, Lynch School of Education, Boston College.

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

Nasjonalt organ for kvalitet i utdanningen (NOKUT). (2008). Evaluering av ingeniørutdanningeni Norge 2008 [2008 Evaluation of the Norwegian Engineer Education]. Oslo, Norway: Author.

Programme for International Student Assessment. (2013). *PISA 2012.* Retrieved from http://www.pisa.no/

Reyna, V. F., & Brainerd, C. J. (2007). The importance of mathematics in health and human judgment: Numeracy, risk communication, and medical decision making. *Learning and Individual Differences, 17*, 147–159. doi: 10.1016/j.lindif.2007.03.010

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. New York, NY: Houghton Mifflin.

Speer, D. C. (1991). Clinically significant change: Jacobson and Truax (1991) revisited. *Journal of Consulting and Clinical Psychology, 60*, 402–408. doi: 10.1037/0022-006X.60.3.402

Trends in International Mathematics and Science Study. (2013). *TIMSS in Norway.* Retrieved 15.10.2013, 2013, from http://www.timss.no/timss05_english.html

Utdanningsdirektoratet. (2014a). *Curriculum for the common core subject of mathematics. Competence aims - competence aims after Year 4*. Retrieved from http://www.udir.no/kl06/MAT1-04/Hele/Kompetansemaal/Kompetansemal-etter-4-arssteget-/?lplang=eng

Utdanningsdirektoratet. (2014b). *Curriculum for the common core subject of mathematics. Competence aims - competence aims after Year 7.* Retrieved from http://www.udir.no/kl06/MAT1-04/Hele/Kompetansemaal/Kompetansemal-etter-7-arssteget/?lplang=eng

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

Wu, H.-H. (1999). Basic skills versus conceptual understanding. A bogus dichotomy in mathematics education. *American Educator, 23*, 1–7.