

The Effects of a Math Racetracks Intervention on the Single-Digit Multiplication Facts Fluency of Four Struggling Elementary School Students

Jennifer Karnes
Anne Barwasser
Matthias Grünke
University Cologne, Germany

There is a significant need for easy-to-implement interventions in the early grades, especially in mathematics, as many children have difficulties in their first years of school with automated basic math facts recall and, therefore, fall behind. Automation of single-digit multiplication tasks represents an important developmental step, as it can neither be accomplished by finger counting nor can more complex operations be mastered without it. Previous research has supported math racetracks as an effective intervention for increasing early math skills of elementary school students. The present study sought to replicate previous findings on the positive effect of a racetrack intervention extended by a multicomponent motivational treatment. Further, we tried to closely comply with the single-case reporting guidelines by Tate et al. (2012) for the purpose of establishing racetracks as evidence-based treatment and continue the work of Lund et al. (2012), Walker et al. (2012), Skarr et al. (2014), and Rivera et al. (2014). Four female elementary-school third graders who faced problems with basic math facts received 9-10 individual training sessions over 7 weeks. Visual analysis of results indicates a level increase in multiplication facts solved correctly from baseline to intervention for all four participants, with an average MBD of 410.45 (range 136.45 to 833.33). Limitations and future directions are discussed.

Keywords: **Automation, Basic Math Facts Fluency, Learning Difficulties, Math Racetracks, DI Flashcards**

Introduction

Significance of Numeracy Skills

Mathematical abilities are vital for students to be successful in school, to handle many daily tasks (e.g., reading the time or managing allowances), and later to be able to find employment (Brown & Snell, 2000; Cihak & Foust, 2008; Watts et al., 2014). Therefore, gaining proficiency in skills like early numeracy or addition and subtraction as well as multiplication and division is critical for

Insights into Learning Disabilities is published by Learning Disabilities Worldwide (LDW). For further information about learning disabilities, LDW's many other publications and membership, please visit our website: www.ldworldwide.org.

achieving the benchmarks for mathematical competence at all grade levels (Mullis et al., 2016; National Mathematics Advisory Panel, 2008; Organization for Economic Co-Operation and Development [OECD], 2014).

Acquisition of elementary skills in this area is a reliable predictor of further mathematical development (e.g., Duncan et al., 2007; Watts et al., 2014). However, as the complexity of mathematical problems in future grades increases, students who do not master basic skills concerning numeracy, addition, subtraction, multiplication, and division have more difficulty mastering those future mathematical concepts and, therefore, are at higher risk of school failure with respect to mathematics (Cai, 2007).

Multiplication Fact Fluency

Fluency in multiplication, including automation and generalization, plays a decisive role in mathematical development as it is the first operation in the higher number space that requires the use of strategies beyond finger counting (Burns et al., 2015; Pólya, 2002; Stein et al., 2006). Grasping the concept of this operation by understanding and interpreting, for example, the symbols 4 x 5 to mean four copies of five, children need time to learn basic facts and automate the basic single-digit multiplication facts recall (Park & Nunes, 2001; Stein et al., 2006; Thornton, 1989). As a fluent arithmetician, a student solves faster by automated recall than by calculating the mathematical fact and using a mental algorithm to find the solution to the task (Lerner, 2003; Logan et al., 1996). Stein et al. (2006) set a benchmark of less than two seconds for an answer to be considered automated. Automation of basic facts recall relieves the burden on working memory with its very limited capacities, thereby allowing the student to concentrate on other components, such as subtasks, story problems, and accuracy (Geary, 2007; Stein et al., 2006; Stood & Jitendra, 2007).

Frequency-Building Procedures

To increase the time students need to complete fundamental multiplication tasks, it is necessary to practice individual items, provide drill and practice besides frequency-building procedures that combine timed repetition with feedback as learning opportunities. Both drill and practice and frequency-building are well-grounded and highly effective strategies (Burns, 2005; De Visscher et al., 2018; Greene et al., 2018). Supporting struggling students in their mathematical development seems to be most successfully performed in one-on-one settings, with a significant amount of effective learning time using non-curriculum-based methods that focus on the current achievement level of the individual student without reference to the competencies the child is expected already to have achieved according to the curriculum (Stevens et al., 2017).

In adding to strategy training, peer tutoring, and schema-based learning, direct instruction and intense drill and practice are evidence-based tools that individually, and combined, have proven particularly helpful and necessary for successfully teaching struggling students basic skills and improving their automation processes (Boon et al., 2019; Butler et al., 2001; Stevens et al., 2017).

Use of Motivational Methods to Increase Student Engagement

Motivational factors influence learning processes and can increase or decrease students' performance. Specifically in the field of mathematics, many children show signs of insecurity concerning math exams or even math courses in their everyday school life (Devine et al., 2018; Hattie, 2009; OECD, 2013). If students experience failure, their motivation to deeply engage with learning decreases, whereas success increases their motivation to engage with further content (e.g., Duhon et al., 2015; Hattie, 2009; Jaffe, 2020). Thus, for students experiencing failure in mathematics, additional motivational factors might help to turn around the natural tendency for motivation to decrease.

Direct instruction (DI) flashcards, positive reinforcement by using students' high scores in addition to visual representation of results in a line diagram and immediate feedback through self-scoring are all well-grounded methods found in current research in different combinations. Combining all those methods into a multicomponent motivational system holds particular promise for increasing the motivation of otherwise struggling students (Jaspers et al., 2017; Montague, 2010; Prater, 2018.

Direct Instruction Flashcards

DI flashcards consist of a predetermined set of targets, such as basic math facts, in a flashcard format. If the student can provide an answer within two seconds, the card is considered processed and given to the student. If not, the card is placed at the third place in the pile of cards so the incorrectly answered question can be presented again quickly (e.g. Hopewell et al., 2011; Skarr et al., 2014).

Immediate Feedback and Positive Reinforcement

Immediate feedback through self-scoring combined with positive reinforcement using high scores and entering scores into a line diagram also are effective ways to improve students' motivation and, therefore, help them tackle learning emotional or subject-specific challenging contents (e.g., Duhon et al., 2015; Eckert et al., 2006; Grays et al., 2017; Van Houten et al., 1974; Wells et al., 2017). This technique enables students themselves to monitor whether a given answer is correct or incorrect and, therefore, to correct errors. When using personal high scores, students compare their personal results with those they achieved the last time they trained. That is, they battle against their own previous results, which decreases anxiety around failure and increases their motivation to perform even better (Duhon et al., 2015). Finally, entering their personal scores into a line diagram shows students visually their individual performance and increases in chronical order, and is easy to interpret by students as well as teachers.

Racetracks

All the above mentioned methods are designed to increase students' motivation to deeply engage with individually challenging learning. To implement them in a playful way, the use of the racetrack method has been documented repeatedly as being enjoyable to students as well as a highly effective direct instruction drill-and-practice and frequency-building method for a variety of mathematical skills (e.g., Beveridge et al., 2005; Erbey et al., 2011; Lund et al., 2012; Pfaff et al., 2013; Skarr et al., 2014; Standish et al., 2012; Walker et al., 2012). In addition to increases in acquisition of the content to be learned, studies detected positive results across settings and participants (i.e., students of varying ages, with and without learning disabilities, attention deficit hyperactivity disorder, emotional and behavioral disorders). Of particular interest to the current study, the research around automation of single-digit multiplication tasks by Lund et al. (2012), Rivera et al. (2014), Skarr et al. (2014), and Walker et al. (2012) focused on automation of multiplication tasks by using racetracks combined with DI flashcards.

Skarr et al. (2014), for example, added DI flashcards to math racetracks for three struggling elementary students, using a single-subject multiple-baseline design across three sets of unknown multiplication facts. During baseline, the students received 15 math facts in random order for measurement. Afterwards, they were presented with the racetrack and the DI flashcards with known math facts to get used to the procedure. On the flashcards 6-7 unmastered and 7-8 mastered math facts were printed. Shown a task, the student had to answer within two seconds by saying the entire statement. If the answer was wrong, the interventionist modeled the correct answer, the student repeated it, and the card was put back in the pile, but three cards back for fast repetition. In addition, the interventionist motivated students by telling them to do their best and praising them if they were able to answer correctly within two seconds.

On the racetrack game board, 28 cells were divided into 14 mastered math facts and 5-7 unknown math facts, which were at least repeated twice on the board. Baseline lasted for 3 to 20 measurement points for a total of 23 days of measurement. According to Skarr et al. (2014), their results clearly indicate the relationship between both applied methods and the mastery of the before unknown math facts.

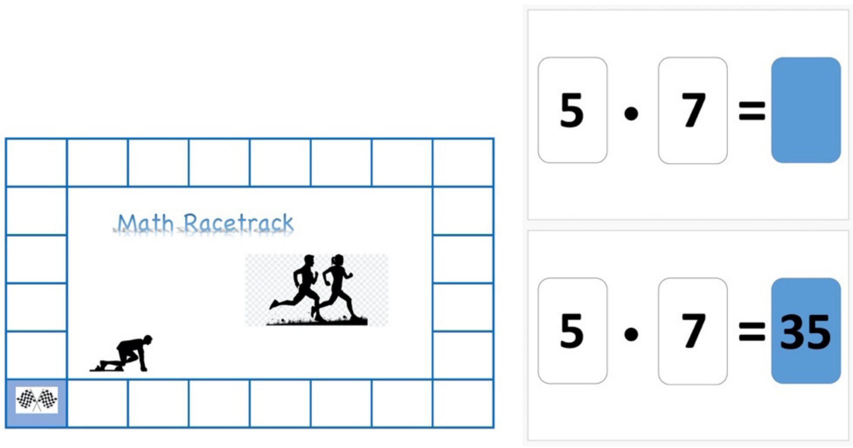

Figure 1. ***Racetrack Game Board and DI Flashcard Example Showing Front and Back***

Using the racetrack method requires a game board that is designed to look like a circuit (e.g., known from Formula 1) with a predefined number of cells. There are many ways to implement a racetrack, such as with a die, by moving from cell to cell, or by using a stopwatch to measure the time needed to complete the racetrack. To play a racetrack, a list of individually unknown math problems or words are written on cards, and each card is placed in one of the cells on the game board. If the student can provide an answer to the card presented within two seconds, then he or she can move on to the next cell or roll the die again, depending on the arranged rules. If the student struggles to answer, corrective feedback is given, and the card is placed back in the cell for the content to be repeated on the next round.

Purpose of the Present Research

Although all the aforementioned publications are single-case experiments reporting medium to large effects and therefore seem to provide a well-grounded technique to foster struggling students, especially concerning their replicability and therefore the methodologically description. Referring to Mulcahy et al. (2016), most research concerning the effects of interventions fostering mathematical skills does not satisfy even the basic standards. This reported lack of high quality interventions is to be considered as well with previous DI and racetrack research.

The Single Case Reporting Guideline in Behavioral Interventions (SCRIBE; Tate et al., 2016), a common research tool to improve the quality of single-case designs, specifies 26 items that should be considered in a single-case study while planning, preparing a manuscript, and reporting results. Items are clustered into the following sections: title and abstract, introduction, method, results, discussion, and documentation. As the centerpiece of research is planning and conducting the methods in detail, while being embedded later into a detailed introduction and a critical interpretation of the results, the reported standards in the method section will be described here in detail, too. To the above cluster, Tate et al. (2016) added a description of the design with all phases and procedural changes, any planned replication and randomization, inclusion criteria for the selection of the participants and their characteristics, ethic approvals, measurement and equipment, intervention, including procedural fidelity, and analysis.

As Horner et al. (2005), Kratochwill et al. (2010), as well as Tate et al. (2016) noted, a set of criteria has been established for determining evidence-based treatments, one of the purposes of this study. Those criteria are as follows: (a) a minimum of five methodologically strong research reports, (b) conducted by at least three different research teams at three different geographical locations, and (c) with the combined number of cases being at least 20.

As the conditions in this study differ from those of Lund et al. (2012), Rivera et al. (2014), Skarr et al. (2014), and Walker et al. (2012), it was not a replication of these studies even though it aimed to add to their findings. The present study was performed in a different setting than the previous research, which was conducted by a research team from the USA, with the aim of meeting more of the aforementioned standards with an additional survey of social validity with respect to the participants.

Referring to successively increasing mathematical complexity in school, on the one hand, and the associated risk of failure correlated with decreasing motivation, on the other hand, the purpose of this research was to evaluate the effects of a math racetrack procedure enhanced by the addition of a multicomponent motivational system consisting of DI flashcards, immediate feedback, and positive reinforcement to address the challenges of four struggling elementary students related to basic single-digit multiplication fact recall. It was hypothesized that the intervention would lead to an overall increase in basic facts computational fluency. Therefore, the aforementioned components in combination were used as the independent variable, whereas the number of single-digit multiplication tasks correctly solved orally within two seconds were defined as the dependent variable.

Method

Participants and Setting

Subjects were six third graders (aged 8 to 9 years old) attending two different classes of an inclusive elementary school with approximately 200 students between grades 1 and 4 in a major city located in North Rhine Westphalia (Germany). The two third-grade classroom teachers proposed 12 students as a preliminary selection of eligible students based on teachers' impression of the students' mathematical competences.

Specifically, the preselection of the students was based on the following inclusion criteria: (a) basic understanding of single-digit multiplication tasks, (b) no to rudimentary automation of single-digit multiplication tasks from 1 x 1 to 10 x 10 according to the teacher, (c) regular school attendance in the previous two months according to the teacher, (d) motivation to be participate in the training according to their own orally given statements, (e) math lessons of the previous two months in class do not include any of the methods of the intervention nor fostering of the automation of single-digit multiplication tasks, and (f) participation in more than 80% of the intervention.

With 12 students meeting the inclusion criteria (a) through (e), we conducted the Multiplication subtest of the Heidelberg Math Test 1-4 (HRT 1-4; Haffner et al., 2005), which was developed to describe computational skills of basic mathematical operations regardless of grade. This instrument was standardized with a calibration sample of $N = 3354$ for the first to fourth grade from 2002-2004. Internal consistency of the HRT 1-4 is $r = -.67$ for math grade, re-test reliability is with $r = .69\text{-}\ .93$. Further, the test can be considered with objectivity concerning implementation, evaluation, and interpretation (Haffner et al., 2005). According to the test manual, students with a percentile between 11^{th} and 25^{th} are specified to be at the borderline of competency, while those between the 0^{th} and 11^{th} percentiles show a marked weakness. To be eligible for the study, students had to reach a percentile between 0^{th} and 25^{th}.

Further, a DIN-A-4 worksheet with 72 single-digit multiplication tasks was administered without a time limit. The items were arranged randomly and excluded tasks with the factors 0, 1, and 10. To be selected for the study, students had to solve at least 50% of the tasks correctly with a processing time $\geq$ 15 minutes to ensure that multiplication as operation was well understood, but the tasks were automated only to a small extent as the aim of this study was to explicitly show how the single-digit multiplication fact recall increases through drill and practice on the basis of a previously established understanding of the operation. To determine how many of the 72 items had already been automated, every student had two seconds to provide an oral answer to the orally and visually presented tasks. According to Stein et al. (2006), an answer given in less than two seconds can be considered to be automated.

Finally, two of the six remaining children had to be excluded later due to the number of days they were not at school during the intervention. Of these two students, data could only be collected at 73% of the measurement points across baseline and intervention. Ultimately, only four female participants met all the criteria (a) to (g). Their names were changed for this study to ensure confidentiality.

The first participant was Anna (9 years old), who was born into a native German-speaking household. Based on the HRT1-4, her multiplication skills showed a "marked weakness," with a percentile of 4 on the Multiplication subtest. Further, she only reached 56% of solved multiplication items (SMI) in 18 minutes, which was the lowest result of all the participants in this study. Anna continued to use her fingers to solve even simple multiplication problems for some tasks. However, she told the teacher and two graduate students who served as interventionists (see below) that she was eager to participate in class. Like all the participants, Anna had no diagnosed disability concerning behavior, learning, or acquisition of mathematical competence.

The second student was Dilara (8 years old), who was of Greek and Turkish migration background but spoke German fluently according to the interventionists. Based on the HRT1-4, she reached a percentile of 8, which was at the "marked weakness" level. On the worksheet with multiplication tasks, she reached 68% SMI in 16 minutes. Dilara claimed to be motivated to participate in the training.

The third participant, Eda (8 years old), also had a Turkish migration background but also spoke German fluently according to the interventionists. She reached a percentile of 10 on the HRT1-4, which was the transition from "marked weakness" to "risk area," She reached greater than 90% SMI on the worksheet after working with her fingers for 33 minutes. She also claimed to be motivated to train with the interventionists.

Betül (8 years old) was also born into a family with a Turkish migration background and spoke fluent German like the other participants. In the Multiplication subtest from the HRT1-4, she reached a percentile of 24 and was able to solve more than 90% of the items from the worksheet. However, like some of the other students, it took her more than 30 minutes. According to the teacher, she needed special attention because she was easily distracted.

For all participating students, the legal guardians or parents were informed by the teacher about the intervention in a personal meeting and a written consent was obtained before the beginning of diagnostic procedure. Further, a contract was drawn up with the school that the study could take place within regular school hours and premises.

Three female graduate university students in special education for children with learning disabilities were chosen as interventionists who conducted the training. They had attended math classes at the university and were close to graduating with a Master of Education for special needs. Prior to the study, they were extensively trained in four 60-minute sessions in a personal lecture-type format by the first author on how to perform the intervention. Additionally, the graduate students, who were experienced in working with children in school settings accompanied with interventions using single-case designs, were instructed to use a detailed script for the implementation. During the study, the interventionists and the first author had at least one phone conversation per week and stayed in contact via email over the entire period of the study.

Dependent Variable and Measurement

To preselect the common intersection of unautomated single-digit multiplication tasks among the participants as the dependent variable for the study, flashcards containing a pool of 72 items (excluding the reversals along with tasks with factor 1) were presented. One of the graduate students presented an item, while another recorded the time required for the answer using a hidden stopwatch to ensure that the participant would not feel pressured to answer quickly but was still motivated to solve a task, even if she exceeded the two-second time limit. The answer was noted on a recording sheet with the following categories: "correct within two seconds," "counted correct (more than two seconds)," "wrong," "wrong with correction," and "no answer." From the pool of 72 tasks, the intersection of the 26 (13 different tasks and their reversals) unknown tasks was filtered out. Those 26 tasks were randomly allocated to prepare 18 record sheets for each measurement over all data point from baseline until maintenance. Further, those record sheets were assigned randomly to each data point per participant to improve the internal validity of measurement.

Experimental Design

A single-subject concurrent multiple-baseline design (AB extension) across participants was applied to evaluate the effects of the training (Horner et al., 2005). In this design, the baseline phase (without treatment) (A) was immediately followed by a treatment phase (with racetracks and the multicomponent motivational system) (B), which consisted of three 20-minute training sessions each followed by measurement. The training was conducted on Monday, Wednesday, and Friday over the entire period of the six-week intervention. To support the demonstration of the long-term effects of the intervention, maintenance data were collected three weeks after the end of the study (Riley-Tillman & Burns, 2009). The treatment onset for each participant was staggered randomly with a baseline phase duration of five to nine days to increase internal validity by controlling history and maturation (Dugard et al., 2012; Tate et al., 2016). This design resulted in Anna and Dilara having 5 probes in Phase A and 10 in Phase B and Eda and Betül having 6 days in the baseline condition and 9 scheduled treatment sessions. For all the participants, Phase E (maintenance without treatment) lasted three days. The two children to have had a baseline length of seven days and an intervention length of eight sessions were excluded due to many missing data. Therefore, implementation of the study underwent procedural changes from the original plan, and only two different baselines can be reported in the results.

Further randomization was used besides allocation of participants to baselines like a randomly allocated set per measurement per student, order of tasks during each training session, treatment integrity observation per student and interventionist, including allocation of interventionists observing treatment integrity.

Procedures

The study was conducted over six weeks with three 20-minute individual sessions each week. Implementation among the three interventionists across the students was alternated to ensure that the effects were independent of the interventionist. For each session, the participants were brought to a separate room by one of the graduate students and led back again after the training.

Baseline Conditions

During measurement, each student was tested by one of the interventionists, who used the flashcards in a prepared order, the corresponding record sheet, and a hidden stopwatch. Feedback was given with constant interjections, independent of the accuracy of the students' answers. To ensure the conditions were comparable to those in the intervention, the students had to read short stories for 15 minutes after they were tested and, therefore, received a similar amount of attention from the interventionist at the same time a comparable effort regarding their ability to concentrate as during Phase B. Texts were chosen that currently were available in the classroom and met students' reading level according to teacher statements.

Treatment

Measurement during the treatment phase occurred the same way as throughout baseline conditions at the beginning of each session. Training sessions occurred in the same separate room and were structured according to a specially designed manual. The instructions included incorporating every item at least once in the beginning of the session using the DI flashcards, followed by a minimum of one round of the racetrack. On the racetrack, participants were presented turn by turn with one of the 26 items. With a correct answer, their figure could move on to the next cell. If the answer was incorrect, immediate feedback was given, and the interventionist and student repeated the correct solution in unison, the student remained in the current cell, and the math problem was moved into the third position in the deck of flashcards to be repeated fairly quickly. To move to the next cell, the next task was taken from the deck and had to be solved correctly. The stopwatch was used to measure time. After finishing the racetrack, students' possibly new high score concerning the time needed to finish the racetrack and the line diagram with the scores from the actual measurement were updated. Further, the student received feedback concerning cooperation during treatment with positive reinforcement through stickers for good working attitude.

Maintenance

After a three-week break following the last training session, maintenance data were collected. In this phase, the same measurement conditions were applied as in the baseline, but without reading after collecting the data.

Materials

In addition to the personal lecture-type format training by the first author, the graduate students received a manual explaining the implementation in detail and created a DIN-A-3 racetrack field with 26 cells along with DIN A-7 flashcards consisting of the identified 26 target tasks, as seen in Figure 1. On the front side of the card, the task was printed, on the backside the task including the solution was printed. The flashcards were used for the training itself as well as the measurement afterwards. For each of the 18 measurement points, a record sheet (see above) was created, and the interventionists used a stopwatch to record reaction time. In addition, a folder was created for each participant containing their personal high score on the front page. The folder was also used to increase the participants' motivation along with a line diagram to record and visualize their improvement and to organize the record sheets in order. Further, the graduate students developed a token system to give the participants feedback concerning their cooperation at the end of each session supplemented with stickers for good cooperation.

A checklist comprising 20 items to consider treatment integrity was created with the categories setting, schedule, materials, procedure, measurement and feedback, dealing with students' behavior as well as space for notes for each session. Finally, a five-item social validity questionnaire was developed to get an impression of the students' acceptance of the intervention.

Treatment Fidelity and Social Validity

The intervention was implemented by the graduate students alternating during the research period so each observed ≥ 35% of the sessions to consider treatment integrity (Noell et al., 2002). The mean correct implementation of the items described in the checklist was 97.3% (range 94 to 100%) across the baseline, intervention, and maintenance conditions. During the interventionists' weekly contact with the first author, steps for improvement of the procedural fidelity could immediately be considered.

After the final training, the interventionists led the students through the social validity questionnaire to get in impression of the acceptance of the intervention by the participants. The interventionists read the questions aloud, and each student answered them orally. The questionnaire consisted of the following statements: “I enjoyed the math racetrack,” “I would like to perform the math racetrack again,” “It was easy for me to remember the tasks,” “I felt comfortable,” “I always look forward to the math racetrack,” and “I enjoyed entering my scores onto the line diagram.” The students could answer each question with “yes,” “a little bit,” or “no.”

Data Analysis

In addition to the descriptive analysis of the graphed data and some prespecified non-overlap indices, a piecewise regression analysis was conducted at the individual level (level 1) as well as across all cases (level 2). For visual analysis, at the same time as statistical analysis, the SCAN package for R (Wilbert & Lueke, 2019) was used to create the plots for each case in addition to computing the non-overlap indices. All effect sizes measurements aim to represent the strength of association between the outcomes and the implemented intervention and, therefore, support the detection of small differences in data between the treatment phases (Vannest & Ninci, 2014). A major argument for using a regression analysis is that only this method adequately takes both level and trend changes into account (Huitema & McKean, 2000).

Results

Table 1 presents the descriptive data of the children for all three phases. The mean scores show the overall improvement between baseline and intervention phases. The maintenance dates show a further increase in data across the participants. As illustrated, Anna displayed the greatest gain in the number of solved multiplication items from baseline to treatment and also in the maintenance sessions, while Dilara showed the least gain when comparing all three phases.

Table 1. Descriptive Data for Each Participant

	n (A)	*n* (B)	*n (E)*	*M (A)* *SD*	*M (B)* *SD*	***M (E) SD***
Dilara	5	10	3	0.80 (1.30)	4.25 (3.33)	5.67 (1.53)
Anna	5	10	3	1.20 (0.84)	11.20 (5.88)	21.00 (1.00)
Betül	6	9	3	4.17 (2.14)	9.86 (0.90)	15.33 (0.58)
Eslem	6	9	3	4.00 (1.90)	13.62 (6.05)	22.33 (0.58)

Note. M = mean, SD = standard deviation, A = Phase A, B = Phase B, E = maintenance, n = measurements.

Figure 2 demonstrates the number of SMIs for each participant. As illustrated, under the treatment conditions, the students' performance generally improved compared to the baseline phase. Phase A for Dilara and Eslem shows a steady baseline, while Betül and Anna show a slightly positive tendency. A remarkable trend can be seen for Anna, which is a decreasing trend in the maintenance measurements. The other probes partly increased after the treatment ended, and some students seemed to be able to continue to advance their progress. Dilara's maintenance measurements display a level effect, and her data are not as stable as the other participants'. In all cases, the variability of the data is quite large, but the improvement is obvious for all four participants.

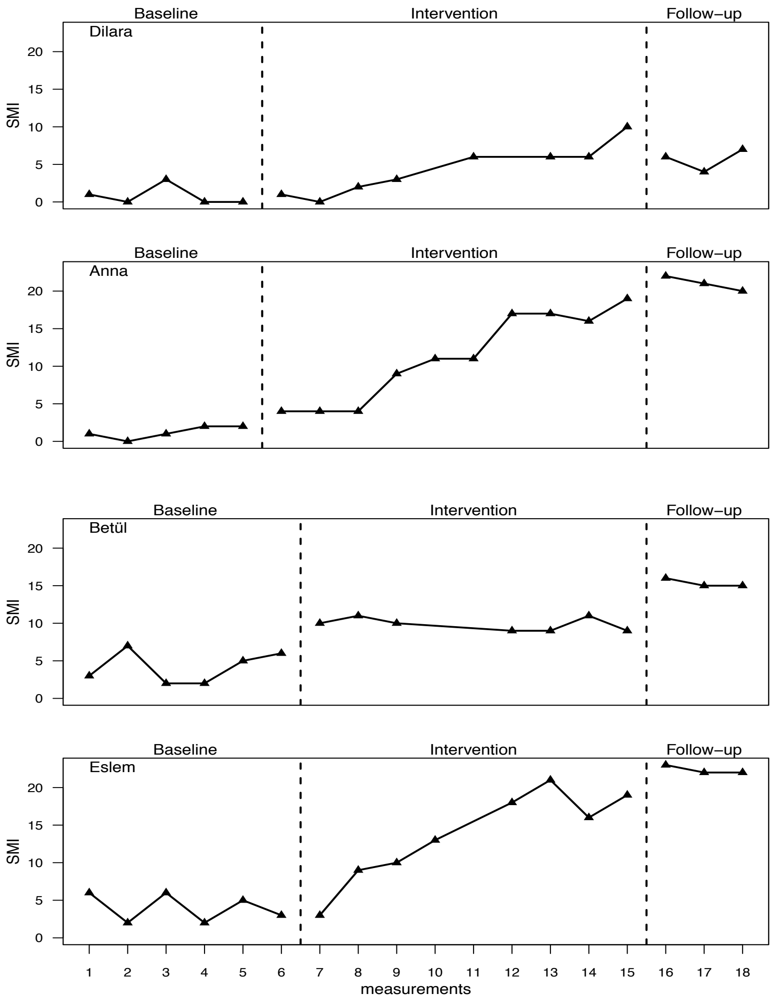

Figure 2. ***Dependent Variable in Phases A, B and Maintenance for Each Participant***

In addition to visual analysis and descriptive data, Table 2 summarizes the results of some non-overlap indices commonly used in single-case research to illustrate the extent of improvement from Phase A to Phase B. The percentage exceeding the median (PEM), non-overlap of all pairs (NAP), Tau-U (Lenz, 2013; Parker et al., 2009; Parker et al., 2011; Vannest & Ninci, 2014), and mean baseline difference (MBD) (Campbell, 2003) were applied. For calculating Tau-U, the option of correcting for the baseline trend (A vs. B + $trend_B$ + $trend_{0A}$) was used.

Table 2. Effect Sizes for the Number of Solved Multiplication Items

	Tau-U	*p*	NAP	*p*	PEM	MBD
Betül	0.32	.090	100	.002**	100	136.45%
Dilara	0.51	.008**	84.00	.03*	87.50	431.25%
Eslem	0.63	.001***	93.00	.005**	87.50	240.75%
Anna	0.76	.000***	100	.001***	100	833.33%

Note. NAP = nonoverlap of all pairs, PEM = percentage exceeding the median, MBD = mean baseline difference. *significant at the .05 level, ** significant at the .01 level, ***significant at the .001 level .

As the students' scores ranged from small/questionable effects to very large effects, the data should be considered more closely. One strength of the PEM is its insensitivity to fluctuations. Values between .70 and <.90 can be considered moderate, and values >.90 can be considered to be strong effects (Ma, 2006). Dilara and Eslem showed moderate effects, while the results for Anna and Betül can be categorized as strong effects.

The NAP compares each data point of one phase with each point of the other phase and, thus, is relatively insensitive to outliers. The benchmarks for moderate effects are .32 to .84; .85-1 can be considered to be a large effect (Parker et al., 2011). While Dilara reached a moderate effect of 84.00 ($p < .05$), Anna (100; $p \leq .001$), Betül (100; $p < .01$), and Eslem (93.00; $p < .01$) achieved large effects.

To underline the effects shown through the non-overlap methods, Tau-U was calculated to represent a significant correlation and differences between Phases A and B. The common benchmarks are .2 to .6 (moderate), .6 to .8 (large), and >.8 (very large) effects (Vannest & Ninci, 2014). While Betül's (.32; $p = .09$) and Dilara's (.51; $p < .01$) results can be considered to be moderate, Eslem's (.63; $p \leq .001$) and Anna's (.76; $p = < .001$) improvements resulted in large effects. Therefore, after implementing the racetrack procedure, all the pupils improved their performance.

Additionally, the mean baseline difference (MBD) was calculated to demonstrate the increase in automated SMIs from the baseline (Campbell, 2003; O'Brien & Repp, 1990). The MBD ranged from 136.45% as the lowest improvement (Betül) to the maximum increase of 833.33% shown by Anna. Nevertheless, Dilara (MBD = 431.25%) and Eda (MBD = 240.75%) also showed remarkable improvement in their performances.

Further, a piecewise regression analysis was conducted at the individual level (level 1, see Table 3) and across all four cases (level 2, see Table 4) (Van den Noortgate & Onghena, 2008). First, no significant baseline trend for any of the students was found. For three of the students, there was a significant slope effect with respect to the comparison between Phases A and B. Dilara showed a significant slope effect ($p<.05$) and improved, on average, by 1.130 in the intervention. Eslem also showed a significant increase ($p<.01$), with an average improvement of 1.205 scale points, with Anna showing similar values, with a significant increase ($p<.05$) and an improvement of 1.467 per intervention session. Only Betül showed no statistical slope effect ($p = .414$), but a statistically significant level effect ($p<.01$) means that there has been a direct improvement from the start of the intervention.

Piecewise regression analysis on the second level reveals a significant slope effect across all participants ($p<.05$); however, no baseline trends were found. In summary, the students increased the number of solved multiplication items by 1.128 per intervention. Thus, the intervention seems to have had a perceivable impact on the dependent variable.

In the social validity questionnaire, all four participants stated unanimously that they enjoyed playing the racetracks and entering their scores on the line diagram. They would also like to continue playing and were looking forward to playing the racetracks again during the intervention phase. Eslem stated that remembering the facts was "a little bit" easy and helped her "a little bit" to solve the tasks given with the flashcards. Anna, Dilara, and Betül gave positive answers to both questions, saying it was easy and it helped in solving the tasks. No one answered "no" to any question, nor did anyone want to add anything to supplement the questions on the questionnaire.

Table 3. **Piecewise regression model for the number of solved multiplication items (Level 1 analysis)**

	B	*SE*	*t*	*p*
		Betül		
Intercept	3.267	1.498	2.180	.054
Trend	0.257	0.385	0.668	.519
Level	5.647	1.700	3.321	.008**
Slope	-0.374	0.438	-0.853	.414
		Dilara		
Intercept	1.400	1.447	0.967	.356
Trend	-0.200	0.436	-0.458	.657
Level	-1.146	1.439	-0.797	.444
Slope	1.130	2.463	2.440	.035*
		Eslem		
Intercept	5.000	2.238	2.235	.047
Trend	-0.286	0.575	-0.497	.629
Level	1.006	2.481	0.405	.693
Slope	2.152	0.653	3.296	.007**
		Anna		
Intercept	0.000	1.524	0.000	1.000
Trend	0.400	0.459	0.871	.401
Level	-1,067	1.501	-0.711	.491
Slope	1.467	0.487	3.015	.011**

*significant at the .05 level, **significant at the .01 level.

Table 4. **Piecewise Regression Model for Number of Solved Multiplication Items (Level 2 Analysis)**

	B	SE	*t*	p
		Overall		
Intercept	2.332	2.258	1.033	.306
Trend	0.052	0.423	0.124	.902
Level	0.932	1.647	0.566	.574
Slope	1.128	0.463	2.438	.018*

*significant at the .05 level.

Discussion

Main Findings

This study examined the effects of a math racetrack game with a multicomponent motivational system, including DI flashcards, immediate feedback through self-scoring and positive reinforcement through visualization of a personal high score, on the single-digit multiplication fact recall of four struggling elementary students. As the results indicate, the treatment increased the participants' learning progress; therefore, it can be considered to be a way of supporting students to overcome their individual challenges.

Previous studies of individual components of the training describe similar conclusions, but in the present research all those components are combined to one intervention that is still easy to implement for teachers in everyday school life and is an attractive way to encourage demotivated learners to deeply engage in learning otherwise fatiguing content. Although the baseline data were not totally stable, the students seemed to have benefited from the intervention to different extents. As the intervention aims to lead demotivated students back on track and to be a tool that is easy for teachers to implement, even with the limitation of variable data, the added value was considerable.

All students' performance improved substantially. Under the baseline conditions, their mean score ranged from 0.80 to 4.17, while during the intervention the range of the mean value was from 4.00 (the lowest) to 15.25 (the highest). The non-overlap indices also support the descriptive analysis showing that the intervention had positive effects on the dependent variable. In addition, the results of the piecewise linear regression analysis of all cases confirm these findings, presenting a statistically significant slope effect in three cases and a statistically significant level effect in one case when comparing Phases A and B. On the second level, a significant slope effect across all participants is displayed by the data, and an increase of 1.13 SMIs per intervention for the students was found.

The maintenance data were collected after a three-week break before the summer holidays. Our data indicate that the effects maintained after the end of the intervention. However, a variability in the data is present, given that Dilara's maintenance data show a level effect while the data of the other students are more stable. Nevertheless, the intervention seemed to improve the performance of all the participants.

Further, responses on the social validity questionnaire give an indication of a high degree of acceptance of the intervention for all students who participated in the study. No one commented negatively on the instruction. Only Eslem answered two questions with more reserve; however, his answer was "a little bit" as opposed to "no."

Overall, the results of this research are compatible with those from the previous studies, all conducted in the USA, by Lund et al. (2012), Rivera et al. (2014), Skarr et al. (2014), and Walker et al. (2012) focusing on the effects of math racetracks on the number of math facts recalled by struggling students with and without disabilities. Therefore, this research adds value in support of designating this easy-to-implement intervention as evidence-based according to the single-case research standards found in SCRIBE (Tate et al., 2016) and establishing a functional relationship (Kazdin, 2010; Ledford & Gast, 2018) between the math racetrack game combined with the multicomponent motivational system.

As reported throughout this paper, this study not only tried to add valuable findings concerning the effectivity of the racetrack intervention itself, but further had the aim to meet as many standards listed in SCRIBE as possible. Clearly, the methods section is the most detailed and the most "vulnerable" part, as this research is taking place under real-life conditions. Still, the design could be clearly identified and described the phases as the phase sequences, procedural changes during the course were noted, the aspects of replication were designated, and randomization, including methods along with the elements that were randomized, are found in the section on the design. Further, the selection of participants is based on the description of inclusion criteria and the method of recruitment. Each participant's demographic characteristics and features relevant to the research question ensuring anonymity are reported as is the setting and location where the study was conducted.

The measurement is explained in detail, and the target behavior and outcome measures, including how and when measurement was applied, are reported. Moreover, the equipment and materials used to measure the target behavior were taken into account. In addition, the intervention itself, baseline condition, and the maintenance are described in sufficient detail to enable replication of the study, and the treatment fidelity is reported in detail, too. The section on the methods concludes with the explanation of the way data analysis was planned and conducted.

The results are reported for each participant, including the score for each session, raw data for the target behavior. Finally, a summary of the findings and interpretation are to be found in the context of current evidence. The concluding section presents the limitations of the study along with applicability and implications.

As mentioned in the beginning of this paper, the impact of motivation on the learning process can both increase and decrease students' performance, and failure in the mathematical development can lead to math anxiety (Devine et al., 2018; Hattie, 2009; OECD, 2013). The findings of this study show how easily struggling students can be engaged into learning contents like automation of single-digit math facts recall if motivational factors are added to instruction that otherwise can have boring and tiring effect on students. As such, the study extends the findings of previous research employing math racetrack combined with DI flashcards and further motivational components.

Limitations

Despite the promising results of the study, a number of limitations deserve mention. First, the third tier in the multiple baseline design is missing as two students had to be excluded due to the fact that they only attended 73% of the measurement time. Although the outcome does not meet the standards of three baseline durations, the procedure itself did so as participants were allocated randomly to the duration of a baseline between five and nine days. Any study that is conducted in a school environment is vulnerable to unintentionally violate best-practice standards, as seen in this case. This underlines the importance of trying to meet current standards while designing further research.

In addition, the sample is too small to draw general conclusions about the effects. However, the findings should not be seen as isolated from previous investigations concerning the effectiveness of math racetracks on the number of automated math facts achieved. On the contrary, one of the main objectives of this study was to substantiate the results obtained to date considering the standards for single-case research and interpreting them in a broader context of previous findings. Additionally, it was not possible for this study to meet all the standards described in Tate et al.'s SCRIBE (2016), such as the absence of interobserver agreement for the measurement and some changes in procedure, which makes it hard to consider the findings as a true replication of previous research. However, many standards were met and allow future research teams to replicate this study due to the precision of the description of all methods and materials. Even with this limitation, this study confirms the effectivity of the intervention.

Another limitation is the lack of data about incorrect responses, which might otherwise have provided a more detailed interpretation of the existing results concerning the correctly solved target tasks. Additionally, the classroom teachers knew about the treatment, and it is possible, therefore, that they unconsciously influenced the intervention and the measurement.

Given that this was a multicomponent motivational intervention, it is not possible to determine with certainty the source of the specific positive effects. To be able to draw conclusions about the change in data due to the exposure to math problems and corrective feedback over baseline and intervention, an alternating-treatments design comparing racetracks and flashcards, for example, would have been helpful. Another limitation is that the interventionists conducted the treatment integrity and performed the measurements, which, according to Podsakoff et al. (2003), might have led to inflated scores.

Finally, the measurement was performed orally using flashcards. The students' time to answer was controlled using a hidden stopwatch. Therefore, the accuracy of the data might not be highly precise, as it leaves room for interpretation. If an answer is correct or goes beyond the two-second time limit that proves automation of knowledge. Using a digital measurement supported by technology, such as PowerPoint, where slides change after two seconds, would have been more accurate. However, the interventionists were instructed by the first author exactly on how to measure the time, control the answers, and complete the protocol sheet; therefore, the results of their measurements can be considered to be comparable.

Practical Implications and Future Research

This investigation provides teachers with further arguments for using this intervention to increase the number of automated mathematical facts among struggling students. Since the results of this study are comparable to the findings obtained by a research team in the USA, it can be seen as progress, and the method can be regarded as slightly more evidence-based. In our case, there were some differences from the previous research concerning the effects of the math racetracks due to greater compliance with the standards. Additionally, the present study included more motivational aspects, such as positive reinforcement through the individual high score, entering scores onto the line diagram, implementing a token system, and providing immediate feedback.

One of the strengths of this economic method is its ease of implementation in everyday school life. Due to its simplicity, peers or parents should be motivated to use this method in addition to teachers. Another strength is the flexibility of the learning content, which can be easily adapted to individual needs by making changes, for example, to the mathematical facts on the flashcards. That is, if a student needs to learn to automate basic addition facts instead of multiplication tasks, this would be simple to change.

Future research should aim to fulfill more of the standards for single-case studies, such as SCRIBE, with the purpose of being able to call the method evidence-based. In addition, further research should evaluate the effects of using math racetracks in a peer tutor setting more closely. It would be interesting to determine whether students would show the same improvement when trained by a classmate. With reference to the multicomponent nature of the intervention, especially regarding the motivational aspects, one could investigate to what extent these aspects influence the course of learning when using racetracks.

Additionally, there is great need for further research concerning digitalized racetracks for struggling students. A digital approach could increase students' motivation and could be easier for a teacher to implement. In their review of literature about digital-based math fluency interventions, Cozard and Riccomini (2016) determined a great need and pointed out that even less research is to be found that meets scientific standards. Developing a digital version of the racetrack should not be too challenging. A digital version of the racetrack could help, for example, when selecting the items to be learned and during the instruction, where the immediate feedback on each math fact would be provided automatically by the program, freeing up the class teacher's time to provide other needed help.

In summary, the racetrack combined with a multicomponent motivational system is simple to implement, and this study showed that the children had fun and considered the intervention useful. Effective methods like this that can be implemented easily and quickly are of great importance in today's school environment to add to teachers' toolbox while supporting student development.

References

Beveridge, B. R., Weber, K., Derby, K., & McLaughlin, T. (2005). The effects of a math racetrack with two elementary students with learning disabilities. *The International Journal of Special Education, 20*(2), 58–65.

Boon, R. T., Urton, K., Grünke, M., & Rux, T. A. (2019). Mnemonic strategies in mathematics instruction for students with learning disabilities: A narrative review. *Learning Disabilities: A Multidisciplinary Journal*, *24*(2), 49–62. https://doi.org/10.18666/LDMJ-2019-V24-I2-9901

Brown, F., & Snell, M. E. (2000). *Instruction of students with severe disabilities.* Prentice Hall.

Burns, M. K. (2005). Using incremental rehearsal to increase fluency of single-digit multiplication facts with children identified as learning disabled in mathematics computation. *Education & Treatment of Children, 28*(3), 237–249. https://doi.org/10.1037/spq0000312

Burns, M. K., Ysseldyke, J., Nelson, P. M., & Kanive, R. (2015). Number of repetitions required to retain single-digit multiplication math facts for elementary students. School Psychology Quarterly, 30(3), 398–405. https://doi.org/10.1037/spq0000097

Butler, F. M., Miller, S. P., Lee, K., & Pierce, T. (2001). Teaching mathematics to students with mild-to-moderate mental retardation: A review of the literature. *Mental Retardation*, *39*(1), 20–31. https://doi.org/10.1352/0047-6765(2001)039<0020:TMTSWM>2.0.CO;2

Cai, J. (2007). Improving mathematics learning: lessons from cross-national studies of Chinese and U.S. students. *Phi Delta Kappan*, *82*(5), 400–404. https://doi.org/10.1177/003172170108200511

Campbell, J. M. (2003). Efficacy of behavioral interventions for reducing problem behavior in persons with autism: A quantitative synthesis of single-subject research. *Research in Developmental Disabilities, 24*(2), 120–138. https://doi.org/10.1016/S0891-4222(03)00014-3

Cihak, D. F., & Foust, J. L. (2008). Comparing number lines and touch points to teach addition facts to students with autism. *Focus on Autism and Other Developmental Disabilities*, *23*(3), 131–137.

Cozad, L. E., & Riccomini, P. J. (2016). Effects of digital-based math fluency interventions on learners with math difficulties: A review of the literature. *The Journal of Special Education Apprenticeship*, *5*(2), Article 2. https://scholarworks.lib.csusb.edu/josea/vol5/iss2/2

Devine, A., Hill, F., Carey, E., & Szücs, D. (2018). Cognitive and emotional math problems largely dissociate: Prevalence of developmental dyscalculia and mathematics anxiety. *Journal of Educational Psychology*, *110*(3), 431–444. https://doi.org/10.1037/edu0000222

De Visscher, A., Vogel, S. E., Reishofer, G., Hassler, E., Koschutnig, K., De Smedt, B., & Grabner, R. H. (2018). Interference and problem size effect in multiplication fact solving: Individual differences in brain activations and arithmetic performance. *NeuroImage, 172,* 718–727. https://doi.org/10.1016/j.neuroimage.2018.01.060

Dugard, P., File, P., & Todman, J. (2012). *Single-case and small-n experimental designs: A practical guide to randomization tests.* Routledge. https://doi.org/10.4324/9780203180938

Duhon, G. J., House, S., Hastings, K., Poncy, B., & Solomon, B. (2015). Adding immediate feedback to explicit timing: An option for enhancing treatment intensity to improve mathematics fluency. *Journal of Behavioral Education, 24*(1), 74–87. https://doi.org/10.1007/s10864-014-9203-y

Duncan, G., Claessens, A., Huston, A., Pagani, L., Engel, M., Sexton, M., Dowsett, C., Magnuson, K., Klebanov, P., Feinstein, L., Brooks-Gunn, J., & Duckworth, K. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428–1446. https://doi.org/10.1037/0012-1649.43.6.1428

Eckert, T. L., Lovett, B. J., Rosenthal, B. D., Jiao, J., Ricci, L. J., & Truckenmiller, A. J. (2006). Class-wide instructional feedback: Improving children's academic skill development. In S. V. Randall (Ed.), *Learning disabilities: New research* (pp. 167–185). Nova Science Publishers.

Erbey, R., McLaughlin, T. F., Derby, K. M., & Everson, M. (2011). The effects of using flashcards with reading racetrack to teach letter sounds, sight words, and math facts to elementary students with learning disabilities. *International Electronic Journal of Elementary Education, 3*(3), 213–226. http://www.iejee.com/index.html

Geary, D. C. (2007). Development of mathematical understanding. In D. Kuhn, R. S. Siegler, W. Damon, & R. M. Lerner (Eds.), *Handbook of child psychology* (pp. 777–810). Wiley. https://doi.org/10.1002/9780470147658.chpsy0218

Grays, S. D., Rhymer, K. N., & Swartzmiller, M. D. (2017). Moderating effects of mathematics anxiety on the effectiveness of explicit timing. *Journal of Behavioral Education, 26,* 188–200. https://doi.org/10.1007/s10864-016-9251-6

Greene, I., McTiernan, A., & Holloway, J. (2018). Cross-age peer tutoring and fluency-based instruction to achieve fluency with mathematics computation skills: A randomized controlled trial. *Journal of Behavioral Education, 27*(1), 145–171. https://doi.org/10.1007/s1086 4-018-9291-1

Haffner, J., Baro, K., Parzer, P., & Resch, F. (2005). *HRT 1-4: Heidelberger Rechentest. Erfassung mathematischer Basiskompetenzen im Grundschulalter* [Heidelberg Arithmetic Test for 1st- to 4th-Graders]. Hogrefe.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement.* Routledge. https://doi.org/10.4324/9780203887332

Hayter, S., Scott, E., McLaughlin, T. F., & Weber, K. P. (2007). The use of a modified direct instruction flashcard system with two high school students with developmental disabilities. *Journal of Physical and Developmental Disabilities, 19*(4), 409–415. https://doi.org/10.1007/s10882-007-9059-3

Hopewell, K., McLaughlin, T. F., & Derby, K. M. (2011). The effects of reading racetracks with direct instruction flashcards and a token system on sight word acquisition for two primary students with severe conduct disorders. *Electronic Journal of Research in Educational Psychology, 9*(24), 693–710. https://doi.org/10.25115/ejrep.v9i24.1462

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single-subject research to identify evidence-based practice in special education. *Council for Exceptional Children, 71*(2), 165–179. https://doi.org/10.1177/001440290507100203

Huitema, B. E., & McKean, J. W. (2000). Design specification issues in time-series intervention models. *Educational and Psychological Measurement, 60*(1), 38–58. https://doi.org/10.1177/00131640021970358

Jaffe, E. (2020). Mindset in the classroom: Changing the way students see themselves in mathematics and beyond. *Clearing House: A Journal of Educational Strategies, Issues and Ideas, 93*(5), 255–263. https://doi.org/10.1080/00098655.2020.1802215

Jaspers, K. E., McCleary, D. F., McCleary, L. N., & Skinner, C. H. (2017). Evidence-based interventions for math disabilities in children and adolescents. In L A. Theodore (Ed.), *Handbook of evidence-based interventions for children and adolescents* (pp. 99–110). Springer.

Kazdin, A. E. (2010). *Single case research designs: Methods for clinical and applied settings* (2nd ed.). Oxford University Press.

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). *Single-case designs technical documentation.* http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf.

Ledford, J. R., & Gast, D. L. (2018). Single case research methodology. Routledge. https://doi.org/10.4324/9781315150666

Lenz, A. S. (2013). Calculating effect size in single-case research: A comparison of non-overlap methods. *Measurement and Evaluation in Counseling and Development, 46*(1), 64–73. https://doi.org/10.1177/0748175612456401

Lerner, J. (2003). *Learning disabilities: Theories, diagnosis, and teaching strategies* (9th ed.). Houghton Mifflin.

Logan, G. D., Taylor, S. E., & Etherton, J. L. (1996). Attention in the acquisition and expression Automaticity. *Journal of Experimental Psychology: Learning Memory and Cognition. 22*(3), 620- 638. http://doi.org/10.1037/0278-7393.22.3.620.

Lund, K., McLaughlin, T. F., & Neyman, J. (2012). The effects of DI flashcards and math racetrack on multiplication facts for two elementary students with learning disabilities. *The Journal of Special Education Apprenticeship, 1*(1), 1–15.

Ma, H. H. (2006). An alternative method for quantitative synthesis of single-subject researches: Percentage of data points exceeding the median. *Behavior Modification, 30*(5), 598–617. https://doi.org/10.1177/0145445504272974

Montague, M. (2010). Mathematics. In R. T. Boon, & V. G. Spencer (Eds.), *Best practices for the inclusive classroom* (pp. 205–223). Prufrock Press.

Mulcahy, C. A., Krezmien, M. P., & Travers, J. (2016). Improving mathematics performance among secondary students with EBD: A methodological review. *Remedial and Special Education, 37*(2), 113–128. https://doi.org/10.1177/0741932515579275

Mullis, I.V.S., Martin, M. O., & Foy, P. (2016). *TIMSS 2015 international results in mathematics.* TIMSS & PIRLS International Study Center, Boston College.

National Mathematics Advisory Panel. (2008). *Foundations for success: The final report of the National Mathematics Advisory Panel.* U.S. Department of Education.

Noell, G. H., Gresham, F. M., & Gansle, K. A. (2002). Does treatment integrity matter? A preliminary investigation instructional implementation and mathematics performance. *Journal of Behavioral Education, 11,* 51–67. https://doi.org/10.1023/A:1014385321849

O'Brien, S., & Repp, A. C. (1990). Reinforcement-based reductive procedures: A review of 20 years of their use with persons with severe or profound retardation. *Journal of the Association for Persons With Severe Handicaps, 15*(3), 148–159. https://doi.org/10.1177/154079699001500307

Organisation for Economic Co-Operation and Development (OECD). (2014, February). *PISA 2012 results: What students know and can do – Student performance in mathematics, reading and science* (Vol. I, rev. ed.). PISA. http://dx.doi. org/10.1787/9789264201118-en

Park, J. H., & Nunes, T. (2001). The development of the concept of multiplication. *Cognitive Development, 16*(3), 763–773. https://doi.org/10.1016/S0885-2014(01)00058-2

Parker, R. I., Vannest, K. J., & Brown, L. (2009). The improvement rate difference for single case research. *Exceptional Children, 75*(2), 135–150. https://doi.org/10.1177/001440290907500201

Parker, R. I., Vannest, K. J., & Davis, J. L. (2011). Effect size in single-case research: a review of nine non-overlap techniques. *Behavior Modification, 35*(4), 303–322. https://doi.org/10.1177/0145445511399147

Pfaff, E., McLaughlin, T. F., Neyman, J., & Everson, M. (2013). The effects of direct instruction flashcards with math racetrack with addition facts for an elementary school student with ADHD. *International Journal of Basic and Applied Science, 2*(1), 124–130. https://doi.org/10.1177/0145445511399147

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Pólya, G. (2002). The goals of mathematical education. *Mathematics Teaching*, *181*, 42–43.

Prater, M. (2018). *Teaching students with high-incidence disabilities: Strategies for diverse classrooms.* Sage.

Riley-Tillman, T. C., & Burns, M. K. (2009). *Evaluating educational interventions: Single-case design for measuring response to intervention.* Guilford.

Rivera, G., Heric, K., Williams, R. L., McLaughlin, T. F., & Johnson, G. (2014). The use of direct instruction multiplication flashcard and a modified math racetrack procedures on mastery of multiplication facts for an at-risk 6th grade girl. *The Open Education Journal*, *7*(1), 7–11. https://doi.org/10.2174/1874920801407010007

Skarr, A., Zielinski, K., Ruwe, K., Sharp, H., Williams, R. L., & McLaughlin, T. F. (2014). The effects of direct instruction flashcard and math racetrack procedures on mastery of basic multiplication facts by three elementary school students. *Education and Treatment of Children, 37*(1), 77–93. https://doi.org/10.1353/etc.2014.0007

Standish, K., McLaughlin, T. F., & Neyman, J. (2012). The effects of direct instruction flashcard on math facts with one elementary student. *Academic Research International, 2*(2), 241–246.

Stein, M., Kinder, D., Silbert, J., & Carnine, D. W. (2006). *Designing effective mathematics instruction: A direct instruction approach.* Pearson Merrill Prentice Hall.

Stevens, E. A., Rodgers, M. A., & Powell, S. R. (2017). Mathematics interventions for upper elementary and secondary students: A meta-analysis of research. *Remedial and Special Education, 39*(6), 327–340. https://doi.org/10.1177/0741932517731887

Stood, S., & Jitendra, A. K. (2007). A comparative analysis of a number sense instruction in reform-based and traditional mathematics textbooks. *Journal of Special Education, 41*(3), 145–157. https://doi.org/10.1177/00224669070410030101

Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., & Vohra, S. (2016). The single-case reporting guideline in behavioural interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology*, *4*(1), 10–31. https://doi.org/10.1037/arc0000027

Thornton, C. (1989). *Basic number facts – Strategies for teaching and learning. Book i: Addition and subtraction. An activity based program.* Document available from Educational Resources Information Center (ED316425)

Van den Noortgate, W., & Onghena, P. (2008). A multilevel meta-analysis of single-subject experimental design studies. *Evidence-Based Communication Assessment and Intervention, 2*(3), 142–151. https://doi.org/10.1080/17489530802505362

Van Houten, R., Morrison, E., Jarvis, R., & MacDonald, M. (1974). The effects of explicit timing and feedback on compositional response rate in elementary school children. *Journal of Applied Behavior Analysis, 7*(4)*,* 547–555. https://doi.org/10.1901/jaba.1974.7-547.

Vannest, K. J., & Ninci, J. (2014). Evaluating intervention effects in single-case research designs. *Journal of Counseling & Development*, *93*(4), 403–411. https://doi.org/10.1002/jcad.12038

Walker, J., McLaughlin, T., & Derby, K. (2012). The effects of flashcards and a math racetrack on multiplication facts for three rural elementary students at-risk for school failure. *Academic Research International*, *2*(3), 406–419.

Watts, T. W., Duncan, G. J., Siegler, R. S., & Davis-Kean, P. E. (2014). What's past is prologue: relations between early mathematics knowledge and high school achievement. *Educational Researcher, 43*(7), 352–360. https://doi.org/10.3102/0013189X14553660

Wells, J. C., Sheehey, P. H., & Sheehey, M. (2017). Using self-monitoring of performance with self-graphing to increase academic productivity in math. *Beyond Behavior, 26*(2)*,* 57–65. https://doi.org/10.1177/1074295617711207

Wilbert, J., & Lueke, T. (2019). *Package 'scan.'* https://cran.r-project.org/web/packages/scan/scan.pdf

Authors' Note

Correspondence concerning this article should be addressed to: Jennifer Karnes, Department of Special Education & Rehabilitation, University of Cologne, Klosterstr. 79b, Cologne, Northrhine-Westfalia, 50931, Germany, Phone: +49-221-47052147, Email: Jennifer.karnes@uni-koeln.de.