ORIGINAL PAPER

Check for
updates

Comparing the Minimum Celeration Line and the Beat Your Personal Best Goal-Setting Approaches During the Mathematical Practice of Students Diagnosed with Autism

Athanasios Vostanis[1] · Ciara Padden[1] · Aoife McTiernan[2] · Peter E. Langdon[3,4]

Abstract

This study compared two goal-setting approaches found in the Precision Teaching literature, namely the minimum celeration line and the beat your personal best during the mathematical practice of three male students diagnosed with autism, aged 8–9. An adapted alternating treatments design with a control condition was embedded in a concurrent multiple baseline across participants design. Each approach was randomly allocated to either the multiplication/division ($\times\div$) table of 18 or 19, while no approach was allocated to the $\times\div$14 table that acted as a control. Instruction utilized number families and consisted of (a) untimed practice, (b) frequency-building, (c) performance criteria, (d) graphing, and (e) a token economy. Upon practice completion, an assessment of maintenance, endurance, stability, and application (MESA) was conducted. Participants improved with both conditions and maintained their performance well, while improvements with the control condition were weak. The beat your personal best approach was highlighted as slightly more effective in terms of average performance and more efficient in terms of timings needed to achieve criterion. No differences were identified in terms of learning rate (i.e., celeration) or performance on the MESA. More research is warranted to identify which goal-setting procedure is more appropriate for students in special education.

Keywords Precision teaching · Standard celeration chart · Frequency-building · Special education · Number family

✉ Athanasios Vostanis
thanos.vostanis@outlook.com

Extended author information available on the last page of the article

🖄 Springer

Introduction

Mathematics is a core academic skill that has been receiving increasing attention in special education (Browder et al. 2008; Hart Barnett and Cleary 2015; King et al. 2016; Lemons et al. 2015). Despite the increase in focus, students in special education are reported to perform below the expected performance standards. Specifically, 50% of students in the fourth grade and 32% of students in the eighth grade, identified as having a disability, were reported to perform at or above the basic standards of performance in the USA (National Center for Educational Statistics 2019a, b). Similar results have been reported in the UK for students identified as having special educational needs. Specifically, only 33% assessed at the end of Key Stage 1 (a phase of school education for children aged 5–7 years old) and 22% assessed at the end of Key Stage 2 (a phase of school education for children aged 7–11 years old) were reported to meet the basic standards of performance (Department for Education 2019a, b).

Therefore, there is a need for evidence-based approaches to help students achieve the expected outcomes of performance. One approach that could prove beneficial is that of Precision Teaching that focuses on precisely defining, measuring, recording, and analyzing behavior change across time (Kubina and Yurich 2012; Evans et al. 2021). Precision Teaching is a system combining different strategies and tactics as part of its framework. First, behavior is pinpointed by using movement cycles that include an action verb and an object (e.g., Writes digit or Says word). Movement cycles are repeatable while also having a discrete beginning and end (Johnston et al. 2019). That way, Precision Teachers can precisely measure each response's occurrence even if it is emitted in high rates. For example, one would score each occurrence of saying a word during an observation period. Second, the modality of instruction is specified by identifying the learning channels utilized. For example, when a student Sees a math fact and Says the answer, the learning channel set is pinpointed as See–Say. Third, goal-setting is employed by gradually increasing performance expectations. Fourth, ultimate performance criteria are specified as a range (e.g., 80–100 correct answers per minute). Fifth, sensitive progress monitoring is conducted through (a) the use of dimensional measures of behavior (e.g., frequency or rate), (b) a standard visual display called the *standard celeration chart,* and (c) behavioral metrics such as *celeration.* The standard celeration chart offers a series of benefits such as straight trend lines and a proportional view of behavior change. Behavioral metrics quantify all aspects of behavior change across time, such as variability, trend, and immediacy of effect (Calkin 2005; Kubina and Yurich 2012). Precision Teaching, as a measurement system, allows practitioners to engage in a data-based evaluation of students' performance change across time and engage in recursive problem solving when learning is unsuccessful (Kubina et al. 2002; Lindsley 1992). In terms of practice, frequency-building to a performance criterion is employed, which involves timed practice and performance feedback (Kubina and Yurich 2012). Frequency-building to a performance criterion has emerged from the field of Precision Teaching; however, it is not an essential component of the system.

In other words, you can do Precision Teaching without necessarily engaging in frequency-building activities.

Precision Teaching and its components have produced positive outcomes in various areas such as reading (Cavallini et al. 2010; Hughes et al. 2007; O'Brien et al. 2018), writing (Datchuk et al. 2015; Datchuk and Kubina 2014) and mathematics (Chiesa and Robertson 2000; McTiernan et al. 2016; Vostanis et al. 2020). One component, however, that has not been evaluated thoroughly is the goal-setting procedures utilized within a Precision Teaching framework. From a behavior analytic point of view, goal-setting procedures operate on two levels. On the antecedent level, they act as an establishing operation, increasing the value of reinforcement (i.e., value-altering effect) and evoking behaviors associated with gaining access to reinforcement (i.e., behavior-altering effect). Specifically, setting goals increases the value of achieving the specified goal, and it also evokes behaviors associated with accessing reinforcement, such as engaging in the task. On a consequent level, goal-setting approaches can act as reinforcement systems. Specifically, if reinforcement is delivered only contingent on meeting specified goals, then goals act as reinforcement criteria. As for goal achievement, it can produce automatic reinforcement, act as a discriminative stimulus for external reinforcement, or both.

Despite their prominent role in Precision Teaching, a review by Doughty et al. (2004) has highlighted that only 22% of the studies reported the reinforcement system they utilized. The most well-documented goal-setting approach in Precision Teaching is the *minimum celeration line* (MCL) or *celeration aim line* used within the model of generative instruction described by Johnson and Street (2013). In that model, two visual displays, the timings chart and the daily per minute chart, are primarily used to guide instruction. Specifically, based on the baseline measures of performance, teachers plot a celeration line on a ×2 trajectory that guides students to increase their performance by doubling it every week. For example, if a student is performing at ten correct responses on Monday, they will be expected to perform at 20 correct responses the following Monday, something that the celeration line would highlight on the chart. The timings chart is used to evaluate within-session performance, and the daily chart is used to evaluate across-sessions performance. Both visual displays are necessary as they are used in combination to help students achieve the expected weekly growth of performance. This approach has been reported to lead to significant gains in academic performance (Johnson and Street 2012) (Fig. 1).

Another goal-setting approach that has been mentioned in the literature is known as *beat your personal best* (BPB) or simply *personal best* (Ginns et al. 2018; Martin and Elliot 2016; Yu and Martin 2014). In this approach, performance expectations are increased by the minimum amount possible, for example, one more response than the previous best score. This approach has also been reported to lead to improved outcomes. In a recent study, Ginns et al. (2018) compared the effects of the personal best procedure versus no goal-setting during a mathematical fluency-building activity with students aged 10–12 years old. The study demonstrated increased gains for students in the goal-setting condition. This approach has been implemented both within and outside the field of Precision Teaching but has been minimally documented in the Precision Teaching literature. Specifically, Brosnan

Example of the minimum celeration line approach on the timings and daily per minute chart
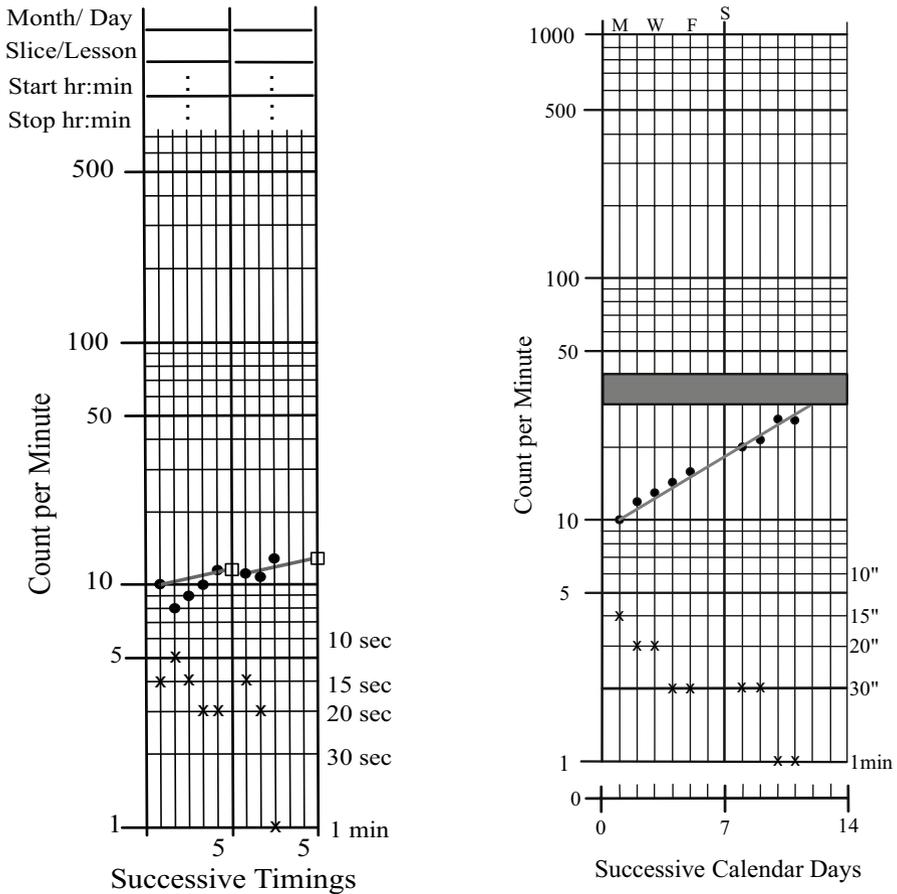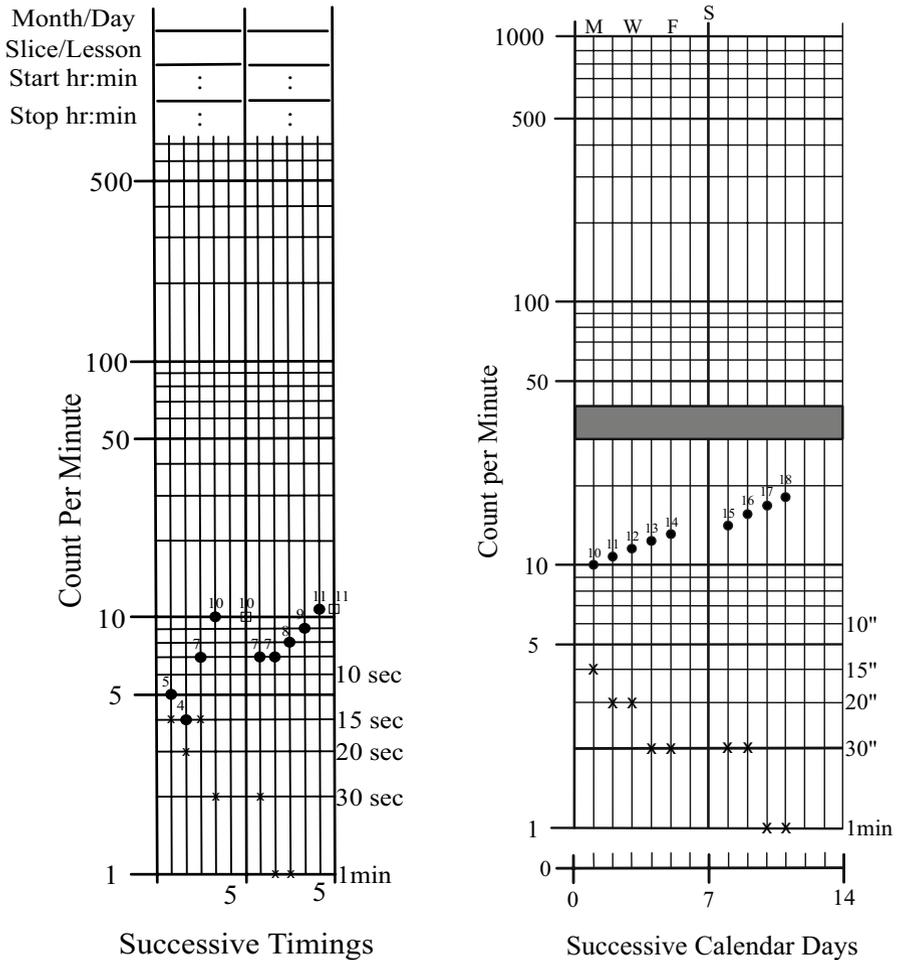

Fig. 1 Example of the minimum celeration line approach on the timings and daily per minute chart. *Note.* The graph on the left is an adapted version of the timings chart. The graph on the right is an adapted version of the daily per minute chart. The dots represent correct responses and the crosses incorrect responses. The gray trend lines represent the minimum celeration line. The open square on the timings chart is the goal box pinpointing the day's performance goal. The gray band on the daily chart shows the ultimate performance criteria set, which are expressed as a range

et al. (2018) conducted a study evaluating Precision Teaching as a Tier 2 intervention for improving foundational reading skills with at-risk kindergarten students. In the study, they used the minimum celeration line as their primary goal-setting approach, but they did also utilize the beat your personal best approach as an additional tactic when the instruction was unsuccessful (Fig. 2).

Examining these two approaches, it is evident that despite their focus on increasing performance, they have two critical differences. First, when using MCL, the focus is upon manipulating learning as the primary variable, while when BPB is

Example of the beat your personal best approach on the timings and daily per minute chart

Fig. 2 Example of the beat your personal best approach on the timings and daily per minute chart. *Note.* The graph on the left is an adapted version of the timings chart. The graph on the right is an adapted version of the daily per minute chart. The dots represent correct responses and the crosses incorrect responses. With the BPB approach, the number of correct responses is written above each datum point and the daily goal is written above the goal box, which is pinpointing the day's performance goal. The gray band on the daily chart shows the ultimate performance criteria set, which are expressed as a range

used, the focus is upon manipulating performance. The distinction between learning and performance is considered critical in Precision Teaching. Learning refers to the student's speed of behavior change across time and is measured through celeration (Lindsley 2000). Performance refers to the student's behavior at a specific point in time, for example, during an examination and is measured through frequency (Lindsley 2000). Second, when using MCL, behavior change across time is calculated using multiplication, for example, from 2 to 8 responses per minute; as such, the

Table 1 Data regarding participants' mathematical abilities, adaptive behavior, and autistic traits

Participants	TEMA-3 Raw scores	TOMA-3 Maths ability index	VABS-II TRF Composite score	GARS-2 Autism index
Alfred	51	81	92	85
Nick	66	106	94	81
Gavin	63	118	87	79

Note: *TEMA-3* Test of early mathematical ability-3rd edition; *TOMA-3* test of mathematical abilities-3rd edition; *VABS-II TRF* Vineland adaptive behavior scales-II teacher rating form; *GARS-2* Gilliam autism rating scale-2nd edition. Standard score on VABS-II excludes the motor domain. TEMA raw scores, instead of standard scores, are reported as participants were above the age range of administration

weekly growth criterion is one that promotes a doubling of performance. When using BPB, behavior change across time is determined using addition, for example, from two to four responses per minute; as such, the criterion is one that promotes a minimum increase relative to the previous best performance. Thus, the two approaches are distinct in their conceptualization and clinical application.

Due to the differences between the two approaches, more information is warranted regarding their application and effectiveness, especially since goal-setting is an essential component of Precision Teaching. To that end, this study aimed to compare the MCL approach and the BPB approach and examine whether they produced different outcomes in terms of participants' performance (i.e., frequency) and learning (i.e., celeration) during mathematical practice.

Methods

Participants and Setting

Three male students with English as their first language participated in the study. Alfred was 9 years and 6 months old; Nick was 8 years and 8 months old; and Gavin was 9 years and 5 months old. Pseudonyms were used. All students had a diagnosis of an autism spectrum disorder (ASD) and were issued an Educational Health Care Plan that provided information about their diagnosis, level of ability, and needs. Alfred and Nick were Caucasian, and Gavin was mixed race.

Participants' performance on two mathematical assessments highlighted that they were of different mathematical abilities (see Table 1). Alfred's score on the Test of Early Mathematics Ability-3 (TEMA-3; Ginsburg and Baroody 2003) suggests that he had mastered basic addition and subtraction, but not basic multiplication (division is not assessed in TEMA-3). What is more, his performance on the Test of Mathematical Abilities-3rd Edition (TOMA-3; Brown et al. 2013) was below average. Nick's performance on TEMA-3 suggests that he had mastered basic addition, subtraction, and multiplication, while his performance on TOMA-3 was average. Finally, Gavin's performance on TEMA-3 also suggests that he had mastered basic addition, subtraction, and multiplication, while his performance on TOMA-3 was

above average. Before the study commenced, participants had received instruction in all basic mathematical procedures, including addition, subtraction, multiplication, and division, as part of their educational provision. They were reported, by their teacher, as fluent with the easier multiplication tables, but not with the more complex ones or with the division tables.

The study was conducted at the participants' school in England, providing special education services for students aged 3–19 years. The curriculum includes self-help, vocational, social, and academic skills. Sessions took place in a 3 × 3 m room, equipped with a camera, a desk, two chairs, and two storage cupboards with all the necessary resources.

Eligibility Criteria

For inclusion in the study, students needed to have (a) a diagnosis of ASD, (b) completed at least 50/72 items of the TEMA-3, (c) participated in at least one week of formal lessons on multiplication and division, (d) not exhibited challenging behavior that would hinder engagement with the instructional procedures. The last three criteria were applied to ensure that students would successfully participate in all the study stages. The study received a favorable ethical opinion from the University of Kent ethics committee. Participants were invited and asked to agree to take part in the study following parental consent to include them.

Materials

Assessment Tools

Along with the TEMA-3, which was used to help determine inclusion, a series of standardized assessments was used descriptively to provide more information on general ability. TEMA-3 is a 72-item test measuring an individual's mathematical ability, including (a) counting proficiency, (b) cardinality, (c) number comparison facility, and (d) elementary arithmetic (Libertus et al. 2013). Its administration lasts approximately 40 min, and the internal consistency has been reported to be between 0.94–0.96 and the test–retest reliability to be between 0.82–0.93 (Ginsburg and Baroody 2003).

The TOMA-3 is a 145-item test assessing (a) mathematical symbols and concepts, (b) computation, (c) mathematics in everyday life, (d) word problems, and (e) the attitude toward maths (supplemental). Its administration lasts approximately 90 min, and internal consistency was reported at 0.96, while test–retest reliability was reported to be 0.89, with the exception of the mathematics for everyday life subtest with 0.73 (Brown et al. 2013). This tool provided additional information on the participants' mathematical abilities.

The Vineland Adaptive Behavior Scales-II Teacher Rating Form (VABS-II TRF; Sparrow et al. 2005) is a 233-item scale measuring adaptive behavior. It includes four domains related to adaptive behavior, namely (a) communication, (b) daily living skills, (c) socialization, and (d) motor skills, and produces an adaptive

behavior composite. Administration lasts approximately 20 min, and standard scores are available for the domain and composite scores. Internal consistency has been reported at 0.98, while test-retest realibility of the adaptive behavior composite has been reported at 0.91 (Sparrow et al. 2005).

The Gilliam Autism Rating Scale-2nd Edition (GARS-2; Gilliam 2006) is a 42-item scale measuring the severity of symptoms related to ASD. It contains three behavioral subscales tailored around the Diagnostic and Statistical Manual of Mental Disorders fourth edition (DSM-IV) and an early developmental history subscale. The four subscales' internal consistency has been reported at 0.94 on average, while test–retest reliability of the total score at 0.88 (Gilliam 2006). VABS-II TRF and GARS-2 provided, respectively, information on the participants' adaptive behavior and symptoms related to ASD.

General Classroom Materials

Participant materials were stored in ring binders sized 21×29.7 cm. Pencils, erasers, notebooks, and digital timers were used. A laminated 'class-shop' catalog, sized 21×29.7 cm, was created with pages in portrait orientation and a 28 Times New Roman font with a picture in the middle of each page, sized 13×15 cm, showing each available item or activity. Finally, a points' board was made in portrait orientation with a Times New Roman 12 font and a 6×6 grid.

A datasheet, a timings graph, and a daily graph were constructed. The datasheet had a 10×10 table divided into five vertical sections each for one day of the week. Each section was divided into two columns (i.e., corrects–incorrects) with five rows, one for each timing. The datasheet also included areas to record the set-criterion timings as well as performance across all relevant testing procedures. Participants were not asked to record how many facts they skipped, to reduce the complexity of the practice; however, we collected data on skipped facts by simply counting their number on each recording sheet (skipped facts are not presented in Fig. 3, to reduce the complexity of the graph).

For the timings graph, the *x*-axis represented the timings completed each day. The axis was divided into five days and allowed participants to graph up to five timings per day. At the end of each week, the graph was replaced with a new one. For the daily graph, the *x*-axis represented school days (i.e., Monday to Friday) and was replaced with a new one every four weeks. Both graphs had a logarithmic *y*-axis. Graphs were constructed based on the timings and daily standard celeration charts (Calkin 2005) but were simplified for ease of use. The datasheets and graphs associated with the MCL approach were always printed in color, and the datasheets and graphs associated with the BPB approach were always printed in black and white, to optimize discrimination between the two approaches.

Materials for Mathematical Practice

All worksheets were created using Microsoft Excel™ and Microsoft Word™. For the untimed practice, we created a laminated page, sized 21×29.7 cm, that had

Participants' Performance with the MCL and BPB Approaches
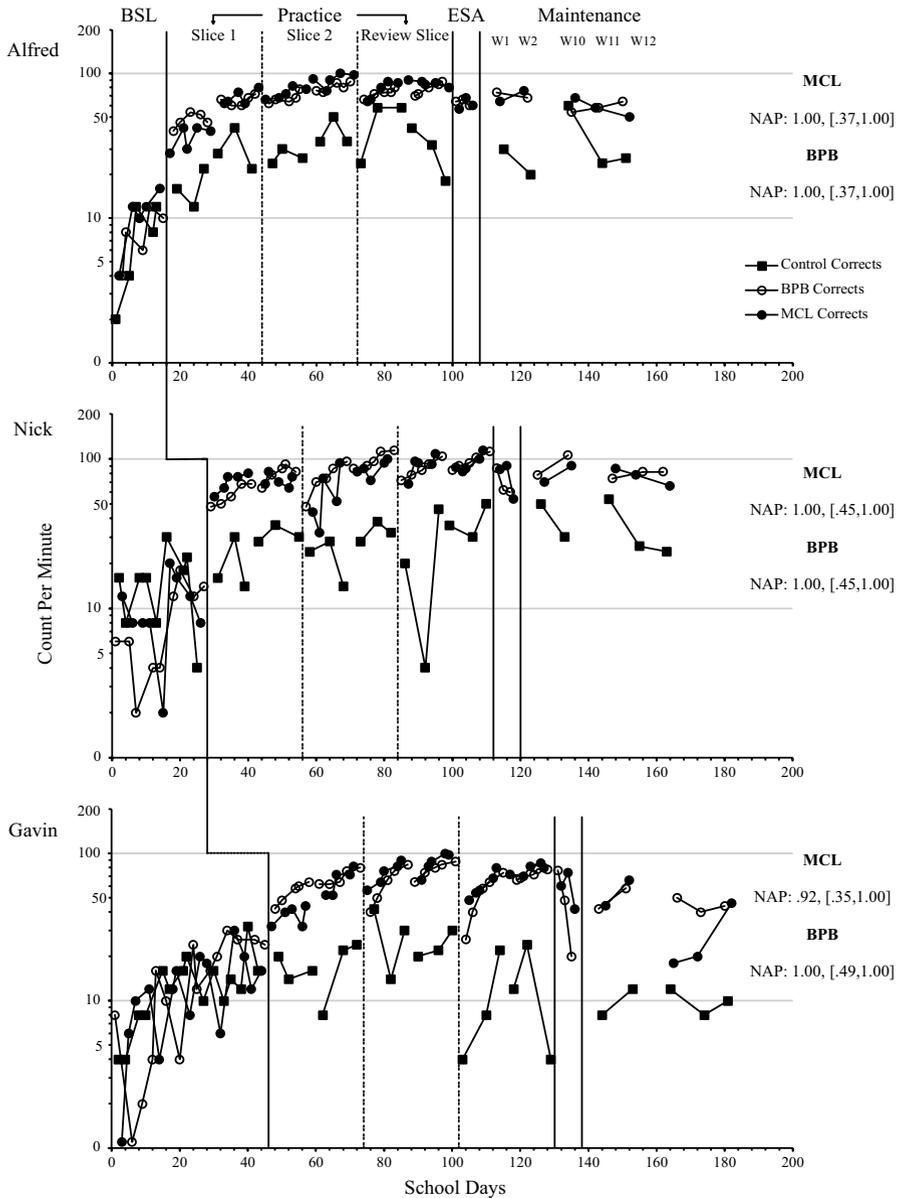
Fig. 3 Participants' performance with the MCL and BPB approaches. *Note.* The control condition is also plotted on the graph. Incorrect digits and skipped facts are not presented to reduce the complexity of the graph. The effect size (NAP) is presented on the right of the graph. Confidence intervals were set at 90%. The assessment of endurance, stability, and application is presented as one condition. Maintenance was assessed after the completion of practice. Data were collected on incorrect responses and skips, but they were not plotted on the graph, to simplify the visual analysis

four 1×5 tables, 5 cm in height, and 11 cm in width. On the top row, we wrote the three numbers that could create a number family (e.g., 18, 2, 36), and the participants wrote all four possible combinations in the remaining rows (i.e., $18 \times 2 = 36$, $2 \times 18 = 36$, $36 \div 2 = 18$, $36 \div 18 = 2$).

For the timed practice and specifically number writing, the worksheet was in landscape orientation and had eight blank rows at a 2.0 distance per page, resembling the lined paper of a notebook, for a total of 15 pages. For the number families, worksheets were in portrait orientation. Multiplication and division facts were aligned to the left and presented horizontally, in an Arial 20 black font, with blank space on the right for participants to write their answers. Each page had ten facts, presented in random order, and each worksheet had 35 pages in total, to ensure that no artificial ceilings would affect participants' performance. Finally, for the Application assessment, a separate worksheet was created through www.themathworkshee tsite.com. That worksheet was in portrait orientation and had 30 multiplication and division facts per page, for a total of ten pages, presented vertically and in random order.

Dependent Variable

Two mathematical skills were assessed: a basic skill (i.e., number writing) and a complex skill (i.e., multiplication/division). Number writing was pinpointed as 'Free[1]-Writes number 0–9 in ascending sequence and with the correct formation on the worksheet.' The dependent variable was the correct and incorrect written digits per minute. Digits were scored as correct if they were written in the appropriate sequence (e.g., 0, 1, 2, 3) and with correct number formation (e.g., fully formed and within the lines). Performance criteria were not set for number writing as it was only assessed. Readers should note, however, that the criterion for that skill is 130–160, correctly formed, digits per minute (Johnson and Street 2013). This skill was assessed as it is an essential skill underlying many mathematical skills.

Multiplication/division was pinpointed as 'See-Writes number of multiplication or division fact presented in random order on the worksheet.' The dependent variable was the correct and incorrect written digits per minute, while we also recorded the skipped facts per minute. Number formation was not assessed for the multiplication/division skill. Performance criteria were set at a frequency of 80–100 correct digits per minute (Johnson and Street 2013). This range was highlighted with a yellow marker on their daily graph so that participants were consistently aware of the expectations in terms of their ultimate performance.

Four multiplication/division tables: $\times \div 13$, $\times \div 14$, $\times \div 18$, $\times \div 19$, were chosen as participants had never practiced them. Out of the four tables, two were ultimately targeted for practice and one acted as a control. Specifically, $\times \div 18$ and $\times \div 19$ were chosen for practice as they were considered of equal difficulty based on the fact that they had an equal number of digits per multiplication fact, while $\times \div 14$ was chosen as

---

[1] Free is another word for Think. We use the word Free, as thinking is free from our observation.

a control based on participants' low baseline performance. To make practice easier for participants, we separated tables 18 and 19 into smaller parts called *slices.* Each table (i.e., $\times\div 18$ and $\times\div 19$) consisted of two slices and a review slice. Slices 1 and 2 included four number families, creating 16 combinations each. Slice 1 included families ranging from $\times\div 2$ to $\times\div 5$ (e.g., $18\times 2=36$ or $90\div 5=18$), and slice 2 included families ranging from $\times\div 6$ to $\times\div 9$ (e.g., $18\times 6=108$ or $144\div 18=8$). Finally, the review slice included all eight number families, creating 32 combinations, and ranging from $\times\div 2$ to $\times\div 9$.

Procedure

Research Design

An adapted alternating treatments design with a control condition (Cariveau et al. 2020) was embedded in a concurrent multiple baseline across participants design (Carr 2005). The order of practice was alternated each day randomly, and the control condition was probed three times a week.

Two goal-setting approaches were compared, namely the MCL approach and the BPB approach. The MCL approach set weekly celeration expectations that participants had to meet. Specifically, participants were expected to double their performance from Monday to Friday following a $\times 2$ celeration. The BPB approach set expectations based on participants' previous best score. In this case, participants were expected to increase their performance by one more digit than their previous best score. We randomly assigned each approach either to the $\times\div 18$ or $\times\div 19$ multiplication/division tables via an online dice roller (https://www.random.org). That way, participants would practice each table with a specific approach. Alfred used the MCL approach with $\times\div 19$ and the BPB approach with $\times\div 18$. Nick and Gavin used the MCL approach with $\times\div 18$ and the BPB approach with $\times\div 19$. Finally, the $\times\div 14$ table was assigned as a control condition with no goal-setting procedure associated with it.

Baseline

During the study, participants did not receive practice on multiplication and division as their teacher focused on other aspects of the curriculum, such as counting, units of measurement, or telling the time. During baseline, participants were provided with one 30 s timing for each skill and were told to perform to their natural pace until they hear the sound of the timer; no instruction or feedback was provided. For number writing, baseline data were collected for five days across two weeks for all participants. For multiplication/division, baseline data were collected in a staggered fashion following the experimental design. Specifically, Alfred's performance was assessed for 5 successive school days, Nick's for 9 days, and Gavin's for 15 days.

Instruction

The lesson was delivered, on a 1:1 format, by an experienced Board-Certified Behavior Analyst (BCBA). Throughout the session, the instructor was present delivering instruction, praise, corrective feedback, and points depending on performance. Corrective feedback during untimed practice included saying the correct answer and asking the participants to write it down before proceeding to the next multiplication/division fact. Corrective feedback during timed practice was provided after the timing was completed in the form of saying the correct answer to the participants. Points were delivered only during the sessions with the instructor, for engaging in untimed practice, timed practice, data collection, and graphing on a variable schedule of reinforcement (VR3). Thus, reinforcement was contingent on engaging with all the practice components. Therefore, in some cases, participants acquired the backup reinforcer despite not having met the performance criterion of the day. That decision was made to keep participants motivated throughout the course of the study. When participants met their daily criterion, they received additional praise and two or three additional points. Overall, participants managed to acquire enough points to access the backup reinforcer in all of the sessions.

For clarity, we will report the common features of both goal-setting approaches and then each one's unique features. At the beginning of each week, participants engaged in two consecutive set-criterion timings that lasted 30 s each. Once both timings were completed, the instructor calculated the performance criteria, and participants started their daily practice that included an untimed element and a timed element. At the early stages of the study, the instructor modeled both the timed and untimed activities and provided additional guidance to students that was faded out over time. During the untimed practice, participants were asked to, simultaneously, write and say all possible multiplication and division combinations of each number family for a total of four families. For example, participants were provided with numbers 18, 2, 36 (which is one number family) and then had to write and say each possible combination. Participants were expected to practice with four number families because slices 1 and 2 had four number families each. In the review slice, where all eight number families were included, participants practiced them in random order. Once they completed a round of untimed practice, they engaged in one 30 s timing and subsequently wrote their correct and incorrect digits on their datasheet and graphed their performance on the timings graph. This process, of untimed and timed practice, was repeated until participants either met their daily criterion or completed five timings. At the end of their practice, participants graphed their best score of the day on the daily graph. Upon completion of their daily practice, participants exchanged their points for a preferred activity or item from the class shop catalog. The catalog included things such as board games, the iPad, Legos, and playing football on the playground. Practice on each slice lasted 10 days for a total of 30 days. The effectiveness of this multicomponent intervention was monitored through the use of Precision Teaching and specifically the use of pinpoints that combined movement cycles and learning channels, as well as the use of the standard celeration chart and behavioral metrics.

Minimum Celeration Line Approach

Despite the common features presented above, each goal-setting approach also had unique features in setting performance criteria and graphing. For the MCL approach, the daily criteria were calculated for the whole week using Microsoft Excel™ based on a ×2 celeration. In terms of graphing, we used the *goal box* and the *minimum celeration line*. We drew the goal box on each day's last line, on the timings graph, to show participants what their daily criterion was (see Fig. 1). Once participants graphed the day's first timing, we connected that datum point with the goal box. That way, participants could see the minimum celeration line, which showed them what their performance's trajectory should be for them to meet their daily criterion. Participants were told that their performance should stay on or above the minimum celeration line. If participants did not meet their daily criterion, they still had to increase their performance to meet the next day's criterion that had already been determined.

Beat Your Personal Best Approach

Contrary to the MCL approach, performance criteria were calculated daily with the BPB approach by increasing the previous day's best score. In terms of graphing, the score of each timing and the goal box was used. Specifically, participants graphed their performance by plotting each datum point on the timings graph and writing their score above it. This approach also used the goal box to show participants their daily criterion. The difference was that no data points were connected to the goal box and that we wrote the criterion number above the goal box. Also, participants wrote their score above each datum point on the daily graph (see Fig. 2). If participants did not meet their criterion for the day, it stayed the same for the next day. That way, we avoided participants dropping their performance on purpose to decrease the next day's criterion.

Assessment of Mastery

When participants completed their practice with the review slice, their performance was assessed for the by-products of fluency through the test of maintenance, endurance, stability, and application (MESA). Following the guidelines from Fabrizio and Moors (2003), endurance was assessed by asking participants to complete a 90 s timing three times longer in duration than their typical timing. Stability was assessed by asking participants to complete a 30 s timing in the presence of distracting stimuli. During this assessment, music played on the iPad, and we also said random numbers to the participants for the whole duration of the timing. The third assessment was that of application. For this assessment, participants completed a 30 s timing with an untaught worksheet, which was in a different format than their typical worksheet. This assessed the application of skills to novel materials. Finally, maintenance was

assessed on weeks 1, 2, 10, 11 and 12 after the practice was concluded. Participants were asked to engage in two 30 s timings, to account for the lack of practice during this phase of the study. That way, participants had the opportunity to engage in a warm-up timing allowing a more accurate evaluation of their performance.

Absence Protocol

From the outset of the study, a protocol was in place to account for any school absence due to illness or other reasons. If participants missed one or two days of practice, then on their return to school, they engaged in one or two double sessions accordingly (e.g., morning and afternoon) to catch up. If participants missed three days of school, then they restarted their weekly practice once they were available. Alfred and Nick did 4 double sessions, and Gavin did 3. The practice was restarted only once for Nick when he was practicing slice 1.

Interobserver Agreement

Interobserver agreement was calculated for all participants and across all phases of the study for $M = 36\%$ (range, 35.5–38%), of the total number of sessions. A BCBA with over ten years of experience independently scored video recordings of the sessions. The agreement was calculated in a two-step manner. First, agreement on correct digits, incorrect digits, and skipped facts was calculated separately by dividing the smaller by the larger number and then multiplying by 100. The three percentages were then added together and divided by three to produce the overall agreement for each skill. This process was repeated for each phase of the study (i.e., baseline, practice, and maintenance). The overall average agreement was calculated by adding the score of each phase and dividing it by three. The average agreement for Alfred was 97% (range, 93–100%), for Nick 88% (range, 82–100%), and for Gavin 94% (range, 87–100%).

Procedural Fidelity

Procedural fidelity was assessed for all participants and across all phases of the study for $M = 36\%$ (range, 35.5–38%), of the total number of sessions, by the same BCBA that collected data on IOA. The baseline checklist included 11 steps, the intervention checklist included 14 steps, and the maintenance checklist included six steps. The intervention checklist included the same number of steps for both the MCL and BPB approaches as both procedures followed the same sequence of untimed practice, timed practice, graphing on the timings graph, and graphing on the daily graph. The BCBA scored each checklist by writing yes or no for each step. Procedural fidelity was 100% across all participants and all phases of the study.

Social Validity

At the end of the study, participants were given a questionnaire (see Appendix) that included 20 open-ended questions about all aspects of their training (e.g., 'how do you feel about graphing your scores?'). Thirteen questions had a scale from 1 to 10 with an unhappy face to the left of number 1 and a happy face to the right of number 10. The happy/unhappy faces were used to help participants discriminate how the scale works. There were also five questions with two options, and students had to circle one of them (e.g., Yes/No or Easy/Hard). Finally, two open-ended questions required an answer from the participants (e.g., what was your favorite part of the practice?). Before participants were left to answer the questionnaire, the instructor said:

There are some questions on this paper about our practice together. I want you to circle you answer. I want you to read the question out loud, and if there is something you did not like, you go toward number 1. The closer you are to number 1, it means that you really did not like something. If there is something you liked, you go toward number 10. The closer you are to number 10, it means that you really liked something. If you circle number five, it means that you did not mind. For some other questions, you will have two choices, and you will need to circle one. Finally, there are some questions that you need to write your own answer.

The instructor was present during the process to provide additional clarification but minimized their interaction to avoid affecting the way participants answered the questions.

Data Analysis

Data were plotted using an online software named PrecisionX, which provided the standard celeration chart for visual analysis and calculated a series of behavioral metrics. PrecisionX was used only by the researchers as the students used paper graphs. Primary metrics utilized were level, celeration, and the level change multiplier. The level shows the average performance of the individual across time. The geometric mean was calculated as it is more appropriate for data plotted on the standard celeration chart, and it is less affected by extreme variables (Clark-Carter 2005). Celeration (i.e., (count/unit of time)/unit of time) is a frequency-derived measure quantifying students' learning rate across time. Celeration can be calculated across days, weeks, months, or even years. In this study, the daily celeration was calculated during baseline and practice, and the weekly celeration was calculated during maintenance as performance was assessed across weeks, not days. The level change multiplier produces a ratio showing how much average performance changed from one phase to another (e.g., baseline to intervention). The ratio was calculated by dividing the highest number by the lowest number and then assigning the multiplication (×) or division (÷) sign to indicate an increase or decrease in average performance across time (Kubina and Yurich 2012). However, all the ratios could be transformed into percentages. For example, a ×2 weekly celeration increase would indicate an increase of 100% per week, while a ÷2 celeration decrease would show

a 50% reduction in performance. For ease of interpretation, all ratios were transformed into percentages.

In addition to these metrics, the Non-Overlap of All Pairs (NAP) was used to calculate the effect of each goal-setting approach on participants' performance. The NAP is an appropriate effect size measure for single-case research with high correlations with the R^2 effect size index (Parker and Vannest 2009). The NAP was calculated only for participants' correct digits by comparing the data from the baseline condition to the data from the maintenance condition. This process was conducted for each goal-setting approach separately. Effect sizes were interpreted following the guidelines by Parker and Vannest (2009). Specifically, weak effects ranged from 0 to 0.65, moderate effects ranged from 0.66 to 0.92, and strong effects ranged from 0.93 to 1.0.

Results

Number Writing

For the basic skill of writing numbers from 0 to 9 with correct number formation, Alfred wrote, on average, 40 correct and 12 incorrect digits per minute across the five days of baseline. Nick wrote 63 correct and 18 incorrect digits per minute, and Gavin wrote 57 correct and 25 incorrect digits per minute. These results show that participants were not performing within the expected range of frequencies.

Multiplication/Division

Duration of Sessions

Participants practiced daily, from Monday to Friday, for a total of 30 days. Alfred practiced for an average of 39 min (range, 22–60 min), Nick practiced for an average of 30 min (range, 15–72 min), and Gavin practiced for an average of 39 min (range, 22–73 min). Session duration included practice with both goal-setting approaches and, where relevant, the control condition. Duration varied across all six weeks as in some cases participants managed to achieve their daily criterion after engaging in one timing, while in other cases, they needed to complete all five timings.

Alfred

During baseline with the MCL approach, Alfred had a mean average of ten correct digits (corrects), 0 incorrect digits (incorrects), and 36 skipped facts (skips) per minute (see Fig. 3, top panel). In the review slice, his performance increased to a mean average of 81 corrects with no incorrects or skips. During maintenance, he had a mean average of 63 corrects with no incorrects or skips. During baseline, corrects accelerated by 596%, incorrects were stable, and skips decelerated by 36%. In the review slice, corrects accelerated by 14% while incorrects and skips were stable.

During maintenance, corrects decelerated by 17%, incorrects accelerated by 48% while skips were stable.

During baseline with the BPB approach, Alfred had a mean average of eight corrects, 0 incorrects, and 26 skips per minute (see Fig. 3, top panel). In the review slice, his performance increased to a mean average of 76 corrects with no incorrects or skips. During maintenance, he had a mean average of 63 corrects with no incorrects or skips. During baseline, corrects accelerated by 379%, incorrects were stable, and skips accelerated by 126%. In the review slice, corrects accelerated by 12%, while incorrects and skips were stable. During maintenance, corrects decelerated by 14% while incorrects and skips were stable.

During baseline with the control condition (i.e., ×÷14), Alfred had a mean average of six corrects, 0 incorrects, and 25 skips per minute (see Fig. 3, top panel). During the weekly assessments, conducted three days per week, his performance increased to a mean average of 29 corrects, two incorrects, and three skips. During maintenance, he had a mean average of 30 corrects, no incorrects, and 13 skips. During baseline, his corrects accelerated by 1900%, his incorrects were stable, and his skips accelerated by 16%. During the weekly assessments, corrects accelerated by 13%, incorrects accelerated by 2%, and skips decelerated by 26%. During maintenance, corrects accelerated by 14%, incorrects accelerated by 66%, and skips decelerated by 60%.

Nick

During baseline with the MCL approach, Nick had a mean average of nine corrects, 0 incorrects, and 22 skips per minute (see Fig. 3, middle panel). In the review slice, his performance increased to 92 corrects with 0 incorrects and 0 skips per minute. During maintenance, he had a mean average of 77 corrects, with 0 incorrects, and 0 skips. During baseline, corrects accelerated by 38%, incorrects decelerated by 19%, and skips accelerated by 44%. In the review slice, corrects accelerated by 11%, incorrects accelerated by 13% while skips were stable. During maintenance, corrects decelerated by 5% while incorrects and skips were stable.

During baseline with the BPB approach, Nick had a mean average of seven corrects, 0 incorrects, and 21 skips per minute (see Fig. 3, middle panel). In the review slice, his performance increased to a mean average of 90 corrects, with no incorrects or skips. During maintenance, he had a mean average of 84 corrects, with no incorrects or skips. During baseline, corrects accelerated by 186%, incorrects decelerated by 25%, and skips accelerated by 50%. In the review slice, corrects accelerated by 20%, while incorrects and skips were stable. During maintenance, corrects decelerated by 8%, incorrects and skips were stable.

During baseline with the control condition, Nick had a mean average of 13 corrects, 0 incorrects, and 15 skips per minute (see Fig. 3, middle panel). During the weekly assessments, he had a mean average of 25 corrects, two incorrect, and three skips. During maintenance, he had a mean average of 35 corrects, two

incorrects, and 0 skips. During baseline, corrects decelerated by 5%, incorrects were stable, while skips accelerated by 46%. During the weekly assessments, corrects accelerated by 6%, incorrects accelerated by 20%, and skips decelerated by 37%. During maintenance, corrects decelerated by 25%, incorrects decelerated by 10%, and skips accelerated by 62%.

Gavin

During baseline with the MCL approach, Gavin had a mean average of ten corrects, two incorrects, and ten skips per minute (see Fig. 3, bottom panel). In the review slice, his performance increased to a mean average of 68 corrects, with no incorrects or skips. During maintenance, he had a mean average of 34 corrects, with two incorrects and two skips. During baseline, Gavin's corrects accelerated by 104%, incorrects accelerated by 9%, while skips decelerated by 6%. In the review slice, corrects accelerated by 35%, while incorrects and skips were stable. During maintenance, corrects decelerated by 39%, incorrects accelerated by 80%, and skips accelerated by 43%.

During baseline with the BBC approach, Gavin had a mean average of ten corrects, 0 incorrects, and 11 skips per minute (see Fig. 3, bottom panel). In the review slice, his performance increased to a mean average of 60 corrects with no incorrects or skips. During maintenance, he had a mean average of 46 corrects with no incorrect digits or skips. During baseline, corrects accelerated by 193%, incorrects decelerated by 27%, and skips decelerated by 32%. In the review slice, corrects accelerated by 62%, incorrects, and skips were stable. During maintenance, corrects decelerated by 8%, incorrects accelerated by 48%, and skips were stable.

During baseline with the control condition, Gavin had a mean average of 12 corrects, 0 incorrects, and 12 skips per minute (see Fig. 3, bottom panel). During the weekly assessments, he had a mean average of 16 corrects, two incorrects, and three skips. During maintenance, he had a mean average of ten corrects, five incorrects, and two skips. During baseline, corrects accelerated by 71%, incorrects decelerated by 23%, and skips decelerated by 11%. During the weekly assessments, corrects decelerated by 10%, incorrects accelerated by 23%, and skips decelerated by 12%. During maintenance, corrects accelerated by 1%, incorrects decelerated by 74%, and skips accelerated by 205%.

Overall Change

With the MCL approach, Alfred's average performance increased by 530% from baseline to maintenance, producing a strong effect size of 1.00 (see Fig. 3, top panel). With the BPB approach, his average performance increased by 688%, producing an effect size of 1.00. With the MCL approach, Nick's average performance increased by 756%, producing an effect size of 1.00. With the BPB approach, his average performance increased by 1100%, producing an effect size of 1.00 (see Fig. 3, middle panel). With the MCL approach, Gavin's performance
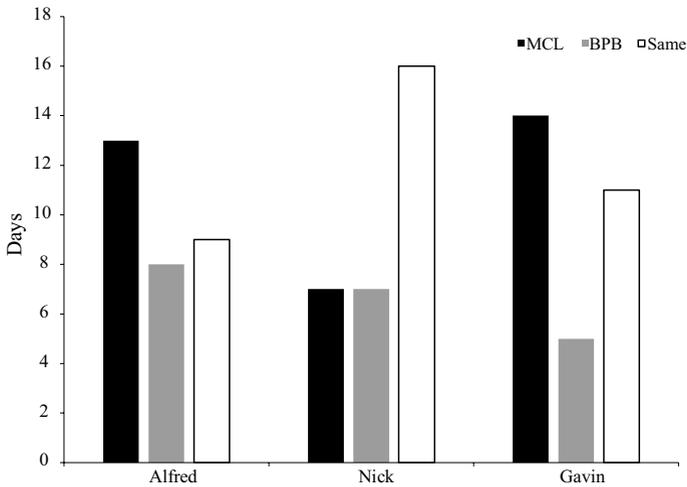
Fig. 4 Number of days participants needed more timings to criterion with MCL or BPB. *Note.* Each white bar shows the number of days where both approaches needed the same number of timings to criterion

increased by 240%, producing an effect size of 0.92 (i.e., moderate). With the BPB approach, his average performance increased by 360%, producing an effect size of 1.00 (see Fig. 3, bottom panel). Examining the effect sizes, we identified a difference only for Gavin as the MCL approach produced a moderate effect size, while the BPB approach produced a strong effect size.

Celeration Values and Number of Timings

To further compare the two approaches, we calculated the average celeration, of correct digits, across the three slices. Alfred's corrects accelerated by 41% (range: 14–84%) with the MCL approach and by 23% (range: 12–36%) with the BPB approach. Nick's corrects accelerated by 28% (range: 7–66%) with the MCL approach and 37% (range: 20–46%) with the BPB approach, Gavin's corrects accelerated by 46% (range: 29–73%) with the MCL approach and 45% (range: 34–62%) with the BPB approach.

To evaluate which approach required participants to engage in more timings-to-criterion, we counted the days that participants needed more timings with one approach over the other (see Fig. 4). For Alfred, the MCL approach required more timings for 13 days, the BPB approach for 8 days, while both approaches required the same number of timings for 9 days. For Nick, both the MCL and the BPB approach required more timings for 7 days, while both approaches required the same number of timings for 16 days. For Gavin, the MCL approach required more timings for 14 days, the BPB approach for 5 days, while both approaches required the same number of timings for 11 days. We also compared the percentage of sessions during which participants completed all five timings. Alfred completed all five timings in 53.30% of the MCL sessions versus 33.33% of the BPB sessions. Nick completed all
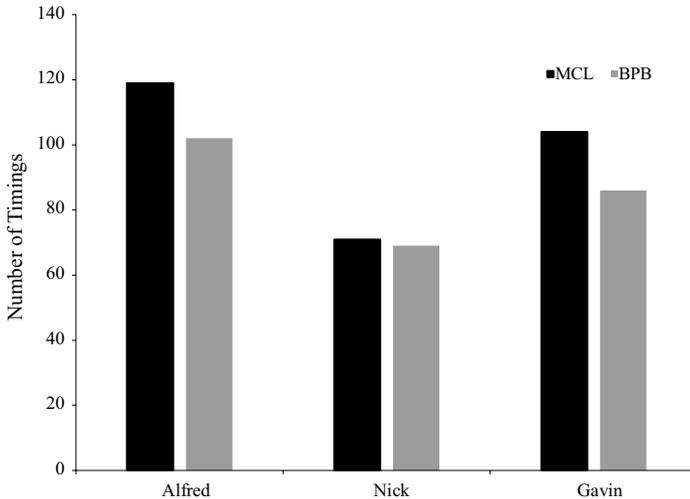
Fig. 5 Total number of timings completed by participants with the MCL and BPB approach

five timings in 10% of the MCL sessions versus 6.66% of the BPB sessions. Gavin completed all five timings in 36.66% of the MCL sessions versus 20% of the BPB sessions. Finally, we calculated the total number of timings completed with each approach (see Fig. 5). Alfred completed 119 timings with the MCL approach versus 102 with the BPB. Nick completed 71 with the MCL approach versus 69 with the BPB. Gavin completed 104 with the MCL approach versus 86 with the BPB.

Comparison of Criteria

The criteria set with each approach were compared to evaluate whether one approach set higher criteria than the other. Alfred had higher criteria with the MCL approach in 63.3% of the practice sessions, and with the BPB approach in 33.3% of the practice sessions, while the criteria were the same in 3.3% of the sessions. Nick had higher criteria with the BPB approach in 60% of the practice sessions, and with the MCL approach in 40% of the practice sessions. Gavin had higher criteria with the MCL approach in 73.33% of the practice sessions, and with the BPB approach in 23.33% of the sessions, while the criteria were the same for 3.33% of the sessions.

We also examined the first and last performance criteria set with each approach. With the MCL approach, Alfred's first criterion was set at ten correct digits per minute, while his final criterion, six weeks later, was set at 70 per minute. With the BPB approach, his first criterion was set at 11 correct per minute, while his final criterion was set at 43 per minute. With the MCL approach, Nick's first criterion was set at 25 correct digits per minute, while his final criterion was set at 54 per minute. With the BPB approach, his first criterion was set at 17 correct per minute, while his final criterion was set at 52 per minute. With the MCL approach, Gavin's first criterion was set at ten correct digits per minute, while his final criterion was set at

52 per minute. With the BPB approach, his first criterion was set at 21 correct digits per minute, while his final criterion was set at 40 per minute.

Finally, we evaluated the number of days, out of a total of 30, that participants met criterion with each goal-setting approach. Alfred met the daily performance criterion in 16 days with the MCL approach and 22 days with the BPB approach. Nick met the daily criterion in 27 days with the MCL approach, and in 28 days with the BPB approach. Finally, Gavin met the daily criterion in 23 days, with the MCL approach, and 29 days with the BPB approach.

Social Validity

Alfred rated the practice as hard; in terms of goal-setting, he preferred the BPB approach, and in terms of graphing, he preferred the MCL approach. When asked which one he would like to use next time, he chose the BPB approach. Nick rated the practice as easy; in terms of goal-setting, he preferred the MCL approach, and in terms of graphing the BPB approach. When asked which one he would like to use next time, he chose the MCL approach. Gavin rated the practice as hard; in terms of goal-setting, he preferred the MCL approach, and in terms of graphing the BPB approach. When asked which one he would like to use next time, he chose the BPB approach.

Discussion

This study aimed to compare two goal-setting approaches that have been utilized in Precision Teaching, namely the minimum celeration line and beat your personal best approach. Students made improvements with high effect sizes with both approaches, while improvements with the control condition were weak. Therefore, this study adds further evidence, as requested by (Ramey et al. 2016), that Precision Teaching can be an effective educational framework for students with developmental disabilities. What is more, the results of the study support the recommendations made by Kubina and Wolfe (2005) on utilizing educational models that focus not only on acquisition but also fluency.

Overall, the differences between the two approaches were minimal in terms of the participants' performance and learning rate. A closer examination, however, suggests that the BPB approach led to a greater overall improvement in performance, as evidenced by the level change multiplier, which was higher for the BPB approach across all participants. This fact is also evident when examining Gavin's effect size results that were slightly better for the BPB approach. Regarding the participants' learning rate (i.e., celeration), the results were mixed. The MCL produced steeper celerations for Alfred, but the BPB produced steeper celerations for Nick while there was no meaningful difference for Gavin. In addition to the previous comparisons, we also examined performance during the assessment of mastery (i.e., MESA), and we did not see any differences between the two conditions. Overall, more research is warranted if we are to understand better how the goal-setting procedures employed in Precision Teaching affect performance and learning. Based on our current results,

it seems that performance with each approach might be affected by idiosyncratic factors. First, a close examination of the social validity outcomes shows that all participants expressed their preference, albeit mixed, in terms of the way their goals were set, and the data were graphed. This fact suggests that preference might have affected participants' performance, a factor that needs to be investigated more in the future. Second, the private verbal behavior produced by participants might have been different with each goal-setting procedure. If participants perceived one approach as more demanding than the other, it could have acted as an establishing operation evoking negative private verbal statements such as 'I will never get it,' which could have subsequently affected performance (Assaz et al. 2018; Friman et al. 1998). During the study, participants did make similar public statements when they considered the goals to be high. Although these public statements were infrequent, they do suggest that this hypothesis might hold merit. This hypothesis is also linked to the number of days that participants reached their daily criterion with each approach. Based on our results, participants met their daily criterion with the BPB approach more often. Although reinforcement was not provided only for reaching the daily performance criterion, it could have affected how participants perceived each approach. In other words, the higher performance criteria set with the MCL approach might have led participants to consider it more difficult than the BPB approach, which would have subsequently affected their private verbal behavior.

Focusing on efficiency, it seems that the BPB approach is the more efficient of the two. Participants needed to engage in fewer timings-to-criterion with the BPB approach. This was evidenced in three ways, first by calculating the number of days where each approach needed more timings that the other, which was lower for the BPB approach for Alfred and Gavin while being exactly the same for Nick. Second, by calculating the percentage of sessions where participants needed to complete all five timings, which was lower for the BPB approach across all participants. Third, by calculating the total number of timings conducted with each approach, which was lower for the BPB approach across all participants. This finding is not surprising considering the difference between the performance criteria set with each approach. Specifically, a comparison of all criteria highlighted that they were higher with the MCL approach for two out of three participants. Also, a comparison between the criteria set on the first day of practice to the criteria set on the last day revealed that they were higher with the MCL approach for all participants. Additionally, the number of days that students met criterion was lower with the MCL approach. In other words, the MCL approach generally set higher criteria for students, which resulted in a lower number of criteria met compared to the BPB approach. This finding is noteworthy as setting higher performance criteria with the MCL approach did not lead to higher performance. On the contrary, the minor differences in overall performance were in favor of the BPB approach.

The way participants' performance increased confirms Lindsley's (1992) observation that behavior increases in a multiplicative manner. In other words, by engaging in frequency-building to a performance criterion, performance is going to multiply in a manner that resembles a logarithmic pattern of growth (i.e., a rapid increase initially and a gradual plateauing of performance as it achieves higher frequencies). Based on our current results, the way performance expectations are set does not

seem to change that pattern. What does seem to be a crucial factor is combining frequency-building with graphing and a systematic increase in performance expectations. This finding also supports the point made by Doughty et al. (2004) about the need to analyze individual components of multicomponent interventions. Considering that the intervention in this study was multicomponent, future research should focus on a component analysis to evaluate each tactic's exact role in a broader precision teaching framework. Specifically, future research should compare engaging in (a) frequency-building, (b) frequency-building plus graphing, (c) frequency-building plus graphing, and goal-setting. Such a comparison would be valuable since precision teachers have historically engaged in multicomponent interventions (Chiesa and Robertson 2000; Johnson and Street 2012; Lin and Kubina 2015; Lokke et al. 2008; Ragnarsdóttir 2007). Generally, more research is needed on goal-setting procedures, a fact already highlighted by Gersten et al. (2009), who conducted a meta-analysis of mathematical interventions and pinpointed goal-setting as a weak component. To that end, it would also be important to further evaluate other goal-setting procedures such as percentile schedules of reinforcement (Athens et al. 2007; Clark et al. 2016; Hall et al. 2009). This approach allows the teacher to set specific response criteria while controlling the density of reinforcement provided to students in a way that resembles the process of shaping (Galbicka 1994). Milyko (2020) has recently provided a detailed account of how percentile (or K-schedules) could be utilized within a Precision Teaching framework. Future research could focus on extending our study by comparing percentile schedules to the minimum celeration line and the beat your personal best approach. It would be beneficial to evaluate whether one of these approaches produces optimized outcomes for students.

Examining participant preference, it seems the BPB approach might be preferred as two participants said that they would like to use that goal-setting procedure in the future. Future research should not only ask participants about their preference but should also follow up by asking participants to engage in additional practice with their preferable goal-setting procedure. That way, we could confirm participants' preferences more robustly.

This study also produced some other findings. First, students with a diagnosis of ASD can benefit from practice using number families, a method that has already produced successful results in mainstream education (McTiernan et al. 2018). Using number families in special education could be highly beneficial as this process 'reduces the amount of memorization needed by three-quarters' (Johnson and Street 2013, p. 142). Second, the number of digits the participants wrote during number writing assessment was lower than the number of digits they wrote while training on multiplication/division. This fact suggests that participants were not performing to the best of their ability, and they would benefit from additional practice on number writing. The fact that they increased their digits while practicing multiplication/division points to spillover effects of the practice to the basic skill. We did not, however, assess number writing in the end to confirm our hypothesis. Nevertheless, this finding highlights the importance of providing students with the opportunity to practice basic skills to fluency, a fact which has already been highlighted in the literature and is particularly relevant for clinical practice (Johnson and Layng 1992; Kostewicz et al. 2019; McDowell and Keenan 2001). By offering more practice opportunities,

students should also contact a dense schedule of reinforcement, which could also be an important variable in increasing their performance.

Limitations

This study had a series of limitations. First, the training we provided in this study was not as comprehensive as it would typically be in clinical practice. We could have assessed and trained a series of basic mathematical skills related to multiplication/division. Second, although participants' performance on number writing was pinpointed as weak, we did not offer training on that skill due to time constraints. Despite the lack of practice on number writing, we do not believe that this affected our ability to answer the research question, which focused on comparing performance between the two goal-setting approaches. Participants practiced with both approaches simultaneously, using multiplication/division tables with the same number of digits, and therefore of equal difficulty, while all other aspects of training remained stable across conditions. Thus, we believe the lack of training on the basic skill did not confound our results. Third, this study had a small number of participants, so the results cannot be generalized to a broader population. Finally, our application assessment focused on the participants' ability to generalize their performance to a novel worksheet. We conducted this assessment following the guidelines for Fabrizio and Moors (2003). Readers should note, however, that application has also been defined as the ability to apply mastered skills when engaging in more complex skills (Kubina and Yurich, 2012; Stocker et al., 2019). We did not assess the latter.

Appendix

The social validity questionnaire provided to the participants at the end of the study.

How do you feel about writing numbers 0-9?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How do you feel about writing x18 and ÷18?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How do you feel about writing x19 and ÷19?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How do you feel about going fast?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How do you feel about using a timer?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How do you feel about maths?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How good are you in maths?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How much have you improved in writing the x18 & ÷18 facts?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How much have you improved in writing the x19 & ÷19 facts?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How do you feel about graphing your scores?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How do you feel about having a goal?

☹   1   2   3   4   5   6   7   8   9   10   ☺

How do you feel about practicing this way?

☹   1   2   3   4   5   6   7   8   9   10   ☺

Which type of setting a goal do you prefer?

Minimum Celeration Line   Beat Best Score

What way of graphing do you prefer?

Beat Best Score   Minimum Celeration Line

Would you like to learn more things this way?

☹   1   2   3   4   5   6   7   8   9   10   ☺

What was the most difficult part of this practice?

What was your favourite part of this practice?

Was this practice easy or hard?

☹ Hard   Easy ☺

Did practicing this way help you?

☺ Yes   No ☹

How would you like to practice next time?

Beat Best Score   Minimum Celeration Line

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval Ethical approval was received from the ethics committee at the University of Kent.

Informed Consent Informed consent was obtained from the participants' parents, and participants' assent was also obtained.

References

Assaz, D. A., Roche, B., Kanter, J. W., & Oshiro, C. K. B. (2018). Cognitive defusion in acceptance and commitment therapy: What are the basic processes of change? *The Psychological Record, 68*(4), 405–418. https://doi.org/10.1007/s40732-017-0254-z.

Athens, E. S., Vollmer, T. R., & Pipkin, C. C. S. P. (2007). Shaping academic task engagement with percentile schedules. *Journal of Applied Behavior Analysis, 40*(3), 475–488. https://doi.org/10.1901/jaba.2007.40-475.

Brosnan, J., Moeyaert, M., Brooks Newsome, K., Healy, O., Heyvaert, M., Onghena, P., & Van den Noortgate, W. (2018). Multilevel analysis of multiple-baseline data evaluating precision teaching as an intervention for improving fluency in foundational reading skills for at risk readers. *Exceptionality, 26*(3), 137–161. https://doi.org/10.1080/09362835.2016.1238378.

Browder, D. M., Spooner, F., Ahlgrim-Delzell, L., Harris, A. A., & Wakemanxya, S. (2008). A meta-analysis on teaching mathematics to students with significant cognitive disabilities. *Exceptional Children, 74*(4), 407–432. https://doi.org/10.1177/001440290807400401.

Brown, V. L., Cronin, M. E., & Bryant, D. P. (2013). *Test of mathematical abilities* (3rd ed.). Austin: PRO-ED.

Calkin, A. B. (2005). Precision teaching: The standard celeration charts. *The Behavior Analyst Today, 6*(4), 207–215. https://doi.org/10.1037/h0100073.

Cariveau, T., Batchelder, S., Ball, S., & La Cruz Montilla, A. (2020). Review of methods to equate target sets in the adapted alternating treatments design. *Behavior Modification*. https://doi.org/10.1177/0145445520903049.

Carr, J. E. (2005). Recommendations for reporting multiple-baseline designs across participants. *Behavioral Interventions, 20*(3), 219–224. https://doi.org/10.1002/bin.191.

Cavallini, F., Berardo, F., & Perini, S. (2010). Mental retardation and reading rate: Effects of precision teaching. *Life Span and Disability, 13*(1), 87–101.

Chiesa, M., & Robertson, A. (2000). Precision teaching and fluency training: Making maths easier for pupils and teachers. *Educational Psychology in Practice, 16*(3), 297–310. https://doi.org/10.1080/713666088.

Clark-Carter, D. (2005). Geometric Mean. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science* (pp. 744–745). John Wiley & Sons.

Clark, A. M., Schmidt, J. D., Mezhoudi, N., & Kahng, S. (2016). Using percentile schedules to increase academic fluency. *Behavioral Interventions, 31*(3), 283–290. https://doi.org/10.1002/bin.1445.

Datchuk, S. M., & Kubina, R. M. (2014). Effects of sentence-combining instruction and frequency building to a performance criterion on adolescents with difficulty constructing sentences. *Journal of Evidence-Based Practices for Schools, 14*(2), 160–185.

Datchuk, S. M., Kubina, R. M., & Mason, L. H. (2015). Effects of sentence instruction and frequency building to a performance criterion on elementary-aged students with behavioral concerns and EBD. *Exceptionality, 23*(1), 34–53. https://doi.org/10.1080/09362835.2014.986604.

Department for Education. (2019a). *National curriculum assessments at key stage 2 in England, 2019 (Revised)*. Retrieved September 1, 2020 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851798/KS2_Revised_publication_text_2019_v3.pdf.

Department for Education. (2019b). *Phonics screening check and key stage 1 assessments in England, 2019*. Retrieved September 1, 2020 from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/851296/Phonics_screening_check_and_key_stage_1_assessments_in_England_2019.pdf.

Doughty, S. S., Chase, P. N., & O'Shields, E. M. (2004). Effects of rate building on fluent performance: A review and commentary. *The Behavior Analyst, 27*(1), 7–23. https://doi.org/10.1007/BF03392086.

Evans, A. L., Bulla, A. J., & Kieta, A. R. (2021). The precision teaching system: A synthesized definition, concept Analysis, and process. *Behavior Analysis in Practice*. https://doi.org/10.1007/s40617-020-00502-2.

Fabrizio, M. A., & Moors, A. L. (2003). Evaluating mastery: Measuring instructional outcomes for children with autism. *European Journal of Behavior Analysis, 4*(1–2), 23–36. https://doi.org/10.1080/15021149.2003.11434213.

Friman, P. C., Hayes, S. C., & Wilson, K. G. (1998). Why behavior analysts should study emotion: The example of anxiety. *Journal of Applied Behavior Analysis, 31*(1), 137–156. https://doi.org/10.1901/jaba.1998.31-137.

Galbicka, G. (1994). Shaping in the 21st century: Moving percentile schedules into applied settings. *Journal of Applied Behavior Analysis, 27*(4), 739–760. https://doi.org/10.1901/jaba.1994.27-739.

Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., & Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Review of Educational Research, 79*(3), 1202–1242. https://doi.org/10.3102/0034654309334431.

Gilliam, J. E. (2006). *Gilliam autism rating scale* (2nd ed.). Austin: Pro-Ed.

Ginns, P., Martin, A. J., Durksen, T. L., Burns, E. C., & Pope, A. (2018). Personal best (PB) goal-setting enhances arithmetical problem-solving. *The Australian Educational Researcher, 45*(4), 533–551. https://doi.org/10.1007/s13384-018-0268-9.

Ginsburg, H. P., & Baroody, A. J. (2003). *Test of early mathematics ability* (3rd ed.). Austin: Pro-Ed.

Hall, S. S., Maynes, N. P., & Reiss, A. L. (2009). Using percentile schedules to increase eye contact in children with Fragile X syndrome. *Journal of Applied Behavior Analysis, 42*(1), 171–176. https://doi.org/10.1901/jaba.2009.42-171.

Hart Barnett, J. E., & Cleary, S. (2015). Review of evidence-based mathematics interventions for students with autism spectrum disorders. *Education and Training in Autism and Developmental Disabilities, 50*(2), 172–185.

Hughes, C. J., Beverley, M., & Whitehead, J. (2007). Using precision teaching to increase the fluency of word reading with problem readers. *European Journal of Behavior Analysis, 8*(2), 221–238. https://doi.org/10.1080/15021149.2007.11434284.

Johnson, K. R., & Layng, T. J. (1992). Breaking the structuralist barrier: Literacy and numeracy with fluency. *American Psychologist, 47*(11), 1475–1490. https://doi.org/10.1037/0003-066X.47.11.1475.

Johnson, K., & Street, E. M. (2012). From the laboratory to the field and back again: Morningside Academy's 32 years of improving students' academic performance. *The Behavior Analyst Today, 13*(1), 20–40. https://doi.org/10.1037/h0100715.

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

Johnston, J. M., Pennypacker, H. S., & Green, G. (2019). *Strategies and tactics of behavioral research and practice* (4th ed.). Abingdon: Routledge.

King, S. A., Lemons, C. J., & Davidson, K. A. (2016). Math interventions for students with autism spectrum disorder: A best-evidence synthesis. *Exceptional Children, 82*(4), 443–462. https://doi.org/10.1177/0014402915625066.

Kostewicz, D. E., Kubina, R. M., & Brennan, K. M. (2019). Improving spelling for at-risk kindergartners through element skill frequency building. *Behavioral Interventions*. https://doi.org/10.1002/bin.1701.

Kubina, R. M., Morrison, R., & Lee, D. L. (2002). Benefits of adding precision teaching to behavioral interventions for students with autism. *Behavioral Interventions, 17*(4), 233–246. https://doi.org/10.1002/bin.122.

Kubina, R. M., & Wolfe, P. (2005). Potential applications of behavioral fluency for students with autism. *Exceptionality, 13*(1), 35–44. https://doi.org/10.1207/s15327035ex1301_5.

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

Lemons, C. J., Powell, S. R., King, S. A., & Davidson, K. A. (2015). Mathematics interventions for children and adolescents with Down syndrome: A research synthesis. *Journal of Intellectual Disability Research, 59*(8), 767–783. https://doi.org/10.1111/jir.12188.

Libertus, M. E., Feigenson, L., & Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learning and Individual Differences, 25,* 126–133. https://doi.org/10.1016/j.lindif.2013.02.001.

Lin, F.-Y., & Kubina, R. M. (2015). Imitation fluency in a student with autism spectrum disorder: An experimental case study. *European Journal of Behavior Analysis, 16*(1), 2–20. https://doi.org/10.1080/15021149.2015.1065637.

Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25*(1), 51–57. https://doi.org/10.1901/jaba.1992.25-51.

Lindsley, O. R. (2000). Celeration and agility for the 2000's. *Journal of Precision Teaching and Celeration, 17*(2), 107–111.

Lokke, G. E. H., Lokke, J. A., & Arntzen, E. (2008). Precision teaching, frequency-building, and ballet dancing. *Journal of Precision Teaching and Celeration, 24*(1), 21–27.

Martin, A. J., & Elliot, A. J. (2016). The role of personal best (PB) and dichotomous achievement goals in students' academic motivation and engagement: A longitudinal investigation. *Educational Psychology, 36*(7), 1285–1302. https://doi.org/10.1080/01443410.2015.1093606.

McDowell, C., & Keenan, M. (2001). Cumulative dysfluency: Still evident in our classrooms, despite what we know. *Journal of Precision Teaching and Celeration, 17*(2), 1–6.

McTiernan, A., Holloway, J., Healy, O., & Hogan, M. (2016). A randomized controlled trial of the Morningside math facts curriculum on fluency, stability, endurance and application outcomes. *Journal of Behavioral Education, 25*(1), 49–68. https://doi.org/10.1007/s10864-015-9227-y.

McTiernan, A., Holloway, J., Leonard, C., & Healy, O. (2018). Employing precision teaching, frequency-building, and the Morningside math facts curriculum to increase fluency with addition and subtraction computation: A randomised-controlled trial. *European Journal of Behavior Analysis, 19*(1), 90–104. https://doi.org/10.1080/15021149.2018.1438338.

Milyko, K. (2020). K-schedules meet precision measurement: A protocol for intervention. *Behavior Analysis in Practice*. https://doi.org/10.1007/s40617-020-00459-2.

National Center for Educational Statistics. (2019a). *The nation's report Card-NAEP mathematic's report card: Grade 4*. https://www.nationsreportcard.gov/mathematics/nation/achievement/?grade=4.

National Center for Educational Statistics. (2019b). *The nation's report Card-NAEP mathematic's report card: Grade 8*. https://www.nationsreportcard.gov/mathematics/nation/achievement/?grade=8.

O'Brien, M., Mc Tiernan, A., & Holloway, J. (2018). Teaching phonics to preschool children with autism using frequency-building and computer-assisted instruction. *Journal of Developmental and Physical Disabilities, 30*(2), 215–237. https://doi.org/10.1007/s10882-017-9581-x.

Parker, R. I., & Vannest, K. (2009). An improved effect size for single-case research: Nonoverlap of all pairs. *Behavior Therapy, 40*(4), 357–367. https://doi.org/10.1016/j.beth.2008.10.006.

Ragnarsdóttir, A. G. (2007). Teaching an Icelandic student with autism to read by combining direct instruction and precision teaching. *European Journal of Behavior Analysis, 8*(2), 289–304. https://doi.org/10.1080/15021149.2007.11434290.

Ramey, D., Lydon, S., Healy, O., McCoy, A., Holloway, J., & Mulhern, T. (2016). A systematic review of the effectiveness of precision teaching for individuals with developmental disabilities. *Review Journal of Autism and Developmental Disorders, 3*(3), 179–195. https://doi.org/10.1007/s40489-016-0075-z.

Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). *Vineland adaptive behavior scales* (2nd ed.). Crowley: AGS Publishing.

Stocker, J. D., Schwartz, R., Kubina, R. M., Kostewicz, D., & Kozloff, M. (2019). Behavioral fluency and mathematics intervention research: A review of the last 20 years. *Behavioral Interventions, 34*(1), 102–117. https://doi.org/10.1002/bin.1649.

Vostanis, A., Padden, C., Chiesa, M., Rizos, K., & Langdon, P. E. (2020). A precision teaching framework for improving mathematical skills of students with intellectual and developmental disabilities. *Journal of Behavioral Education*. https://doi.org/10.1007/s10864-020-09394-2.

Yu, K., & Martin, A. J. (2014). Personal best (PB) and 'classic' achievement goals in the Chinese context: Their role in predicting academic motivation, engagement and buoyancy. *Educational Psychology, 34*(5), 635–658. https://doi.org/10.1080/01443410.2014.895297.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Athanasios Vostanis[1] · **Ciara Padden[1]** · **Aoife McTiernan[2]** · **Peter E. Langdon[3,4]**

[1] Tizard Centre, University of Kent, Cornwallis North East, Canterbury, CT2 7NF, Kent, UK

[2] School of Psychology, National University of Ireland, Galway, Ireland

[3] Centre for Educational Development, Appraisal and Research, University of Warwick, Coventry CV4 8UW, UK

[4] Coventry and Warwickshire Partnership NHS Trust, Coventry CV6 6NY, UK