



## Foster, Colin (2017) Developing mathematical fluency: comparing exercises and rich tasks. Educational Studies in Mathematics . pp. 1-21. ISSN 1573-0816

### Access from the University of Nottingham repository:

<http://eprints.nottingham.ac.uk/46887/1/10.1007%252Fs10649-017-9788-x.pdf>

### Copyright and reuse:

The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the Creative Commons Attribution licence and may be reused according to the conditions of the licence. For more details see:  
<http://creativecommons.org/licenses/by/2.5/>

### A note on versions:

The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact [eprints@nottingham.ac.uk](mailto:eprints@nottingham.ac.uk)

# Developing mathematical fluency: comparing exercises and rich tasks

Colin Foster<sup>1</sup> 

© The Author(s) 2017. This article is an open access publication

**Abstract** Achieving fluency in important mathematical procedures is fundamental to students' mathematical development. The usual way to develop procedural fluency is to practise repetitive exercises, but is this the only effective way? This paper reports three quasi-experimental studies carried out in a total of 11 secondary schools involving altogether 528 students aged 12–15. In each study, parallel classes were taught the same mathematical procedure before one class undertook traditional exercises while the other worked on a “mathematical etude” (Foster *International Journal of Mathematical Education in Science and Technology*, 44(5), 765–774, 2013b), designed to be a richer task involving extensive opportunities for practice of the relevant procedure. Bayesian *t* tests on the gain scores between pre- and post-tests in each study provided evidence of no difference between the two conditions. A Bayesian meta-analysis of the three studies gave a combined Bayes factor of 5.83, constituting “substantial” evidence (Jeffreys, 1961) in favour of the null hypothesis that etudes and exercises were equally effective, relative to the alternative hypothesis that they were not. These data support the conclusion that the mathematical etudes trialled are comparable to traditional exercises in their effects on procedural fluency. This could make etudes a viable alternative to exercises, since they offer the possibility of richer, more creative problem-solving activity, with comparable effectiveness in developing procedural fluency.

**Keywords** Bayesian hypothesis testing · Enlargements · Linear equations · Fluency · Practice · Mastery · Mathematics education · Procedures · Rich tasks

## 1 Introduction

Attaining fluency in key mathematical procedures is essential to students' mathematical development (Department for Education [DfE], 2013; National Council of Teachers of

---

✉ Colin Foster  
colin.foster@nottingham.ac.uk

<sup>1</sup> Centre for Research in Mathematics Education, School of Education, University of Nottingham, Nottingham NG8 1BB, UK

Mathematics [NCTM], 2014; Truss, 2013). Being secure with important mathematical procedures gives students increased power to tackle more complicated mathematics at a more conceptual level (Coddington, Burns, & Lukito, 2011; Foster, 2013b, 2016), since automating skills frees up mental capacity for being creative (Lemov, Woolway, & Yezzi, 2012, p. 36). Devising ways to support the development of robust fluency with mathematical procedures is currently a focus of attention. For example, in England the national curriculum for mathematics emphasises procedural fluency as the first stated aim (DfE, 2013), and the current “mastery” agenda stresses “intelligent practice” as a route to simultaneously developing procedural fluency and conceptual understanding (Hodgen, 2015; National Association of Mathematics Advisors [NAMA], 2016; National Centre for Excellence in the Teaching of Mathematics [NCETM], 2016).

However, a focus on procedural fluency is sometimes seen as a threat to reform approaches to the learning of mathematics, which emphasise sense making through engagement with rich problem-solving tasks (Advisory Committee on Mathematics Education [ACME], 2012; Office for Standards in Education [Ofsted], 2012). In a technological age, in which calculators and computers can perform mathematical procedures quickly and accurately, it may be argued that teaching problem solving should be prioritised over practising procedures. It may also be that an excessive focus on basic procedures fails to kindle students’ interest in mathematics and could be linked to students, especially girls, not choosing to pursue mathematics beyond a compulsory phase (Boaler, 2002). Nevertheless, in a high-stakes assessment culture, where procedural skills are perceived to be the most straightforward ones to assess, the backwash effect of examinations is likely to lead to schools and teachers feeling constrained to prioritise the development of procedural fluency over these other aspects of learning mathematics (Foster, 2013c; Ofsted, 2012; Taleporos, 2005).

In this context, it has been suggested that a mathematics task genre of *etudes* might be capable of addressing procedural fluency at the same time as offering a richer experience of learning mathematics (Foster, 2013b, 2014). Etudes are mathematics tasks in which extensive practice of a well-defined mathematical procedure is embedded within a rich mathematical problem (Foster, 2013b). Such tasks aim to generate plentiful practice incidentally as learners tackle a rich, open-ended problem. East Asian countries which perform well in large-scale international assessments such as the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS) are thought to succeed in emphasising mathematical fluency without resorting to low-level rote learning of procedures (Askew et al., 2010; Fan & Bokhove, 2014; Leung, 2014).

There have been many attempts to design tasks that incorporate meaningful practice (Kling & Bay-Williams, 2015) or exploit systematic variation (NAMA, 2016) to address fluency goals within deeper and more thought-provoking contexts. Not only might this lead to greater interest and motivation for students (Li, 1999), it is conceivable that it could assist in the development of procedural fluency by to some extent shifting students’ focus away from the details of the procedure, perhaps thereby aiding automation. From the point of view of being economical with students’ learning time, Hewitt (1996) described the generation of purposeful practice by subordinating the role of practice to a component of a larger mathematical problem. In this way, attention is placed not on the procedure being performed but instead on the *effect* of its use on a desired goal (Hewitt, 2015).

Mathematical etudes draw on these intentions to situate procedural practice within rich, problem-solving tasks. Although anecdotally etudes have been very favourably received by mathematics teachers, and appear to be popular with students, it is not known whether or not they are as effective as traditional exercises at developing procedural fluency. While etudes

might be expected to offer other advantages, such as greater engagement and opportunities for creative problem solving and exploration, it is not known whether this comes at a cost of effectiveness in narrow terms of developing procedural fluency. It seems possible that diverting students' attention away from the details of carrying out a procedure and onto some wider mathematics problem could hinder their immediate progress in procedural fluency. However, on the contrary, problem-solving aspects of an etude could potentially focus students on the details of a procedure in a way that *supports* development of fluency. So this paper investigates whether or not etudes are as effective as exercises for developing students' procedural fluency.

## 2 Mathematical etudes

### 2.1 Background

Procedural fluency involves knowing when and how to apply a procedure and being able to perform it “accurately, efficiently, and flexibly” (NCTM, 2014, p. 1). The *Mathematical Etudes Project*<sup>1</sup> aims to devise creative ways to help learners of mathematics develop their fluency in important mathematical procedures. It might be supposed that any varied diet of rich problem-solving tasks would automatically generate plentiful opportunities for students to gain practice in a multitude of important mathematical procedures, and that this would be a natural way for procedural fluency to be addressed in the curriculum. However, a rich, open-ended task may be approached in a variety of ways (Yeo, 2017), and, where a choice of approaches is possible, students may be drawn to those which utilise skills with which they are already familiar and comfortable. In this way, areas of weakness may remain unaddressed. For example, a student lacking confidence with algebra may be able to solve a mathematical problem successfully using numerical trial and improvement approaches, or perhaps by drawing an accurate graph. From the point of view of problem solving, selecting to use tools with which one is already competent is an entirely appropriate strategy, but if algebraic objectives were central to why the teacher selected the task, then the task has failed pedagogically. In this way, an open-ended task cannot necessarily be relied on to focus students' attention onto a specific mathematical procedure. Even if it does succeed in doing this, it may not generate sufficient practice of the particular technique to develop the desired fluency, since a broader problem is likely to contain other aspects which also demand the student's time and attention.

For this reason, an etude cannot simply be a problem which provides an *opportunity* for students to use the desired procedure; it must place that procedure at the centre of the students' activity and force its repeated use. Success with the task must be contingent on repeated accurate application of the desired procedure. The *Mathematical Etudes Project* has developed numerous practical classroom tasks which embed extensive practice of single specified mathematical procedures within rich problem-solving contexts (Foster, 2011, 2012a, b, 2013a, b, d, 2014, 2015a). It is whether such tasks are as effective as traditional exercises in developing fluency or not that is the subject of this research.

The term “etude” is borrowed from music, where an etude is “originally a study or technical exercise, later a complete and musically intelligible composition exploring a particular technical problem in an esthetically satisfying manner” (Encyclopaedia Britannica, 2007). Originally, etudes were intended for private practice, rather than performance, but later ones sought

---

<sup>1</sup> [www.mathematicaletudes.com](http://www.mathematicaletudes.com)

to achieve the twin objectives of satisfying an audience in concert as well as working as an effective tool for the development of the performer's fluency. This latter sense inspires the idea of a mathematical etude, which is defined as a mathematics task that embeds "extensive practice of a well-defined mathematical technique within a richer, more aesthetically pleasing mathematical context" (Foster, 2013b, p. 766). In musical etudes, such as those by Chopin, the self-imposed constraint of focusing on (normally) a single specific technique may contribute to the beauty of the music.

The idea of practising a basic skill in the context of more advanced skills is common in areas such as sport (Willingham, 2009, p. 125), and has been used within mathematics education. For example, Andrews (2002) outlined "a means by which practice could be embedded within a more meaningful and mathematically coherent activity" (p. 16). Boaler advised that it is best to "learn number facts and number sense through engaging activities that focus on mathematical understanding rather than rote memorization" (Boaler, 2015, p. 6), and many have argued that algorithms do not necessarily have to be learned in a rote fashion (Fan & Bokhove, 2014). Watson and De Geest (2014) described systematic variation of tasks for the development of fluency, and it is known that, to be effective, practice must be purposeful, and systematically focused on small elements, and that feedback is essential (Ericsson & Pool, 2016). The challenge is to devise mathematics tasks which do this within a rich context.

The three etudes trialled in the studies described in this paper will now be discussed. Two of these etudes address solving linear equations in which the unknown quantity appears on both sides (studies 1 and 2), and the third etude concerns performing an enlargement of a given shape on a squared grid with a specified positive integer scale factor (study 3).

## 2.2 Linear equations etudes

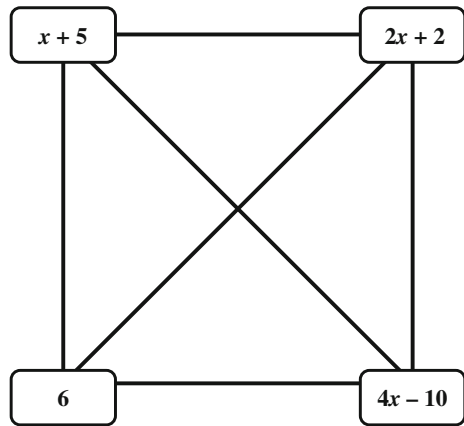
The first two etudes described focus on solving linear equations. Both are intended to generate practice at solving linear equations in which the unknown quantity appears on both sides.

### 2.2.1 Expression polygons etude

In this etude, students are presented with the diagram shown in Fig. 1, called an *expression polygon* (Foster, 2012a, 2013a, 2014, 2015a). Each line joining two expressions indicates that they are equated, and the initial task for students is to solve the six equations produced, writing each solution next to the appropriate line. For example, the top horizontal line joining  $x + 5$  to  $2x + 2$  generates the equation  $x + 5 = 2x + 2$ , the solution to which is  $x = 3$ , so students write 3 next to this line. In addition to recording their solutions on the expression polygon, a student could write out their step-by-step methods on a separate piece of paper.

Having completed this, the students will obtain the solutions 1, 2, 3, 4, 5 and 6. The pattern is provocative, and students typically comment on it (Foster, 2012a, 2015a). This leads naturally to a challenge: "Can you make up an expression polygon of your own that has a nice, neat set of solutions?" Students make choices over what they regard as "nice" and "neat". They might choose to aim for the first six even numbers, first six prime numbers, first six squares or some other significant set of six numbers. Regardless of the specific target numbers chosen, the experimentation involved in producing their expression polygon is intended to generate extensive practice in solving linear equations. Working backwards from the desired solution to a possible equation, and modifying the numbers to make it work, necessitates unpicking the equation-solving process, which could contribute to understanding of and

**Fig. 1** Expression polygons etude. (Taken from Foster, 2015a)



facility with the procedure. Students are expected to attend more to the solutions obtained than they would when working through traditional exercises, where the answers typically form no pattern and are of no wider significance than that individual question. As students gain facility in solving equations, they focus their attention increasingly on strategic decisions about which expressions to choose. They might even go on to explore what sets of six numbers may be the solutions of an expression polygon, or experiment with having five expressions rather than four, for instance. In this way, the task is intended to self-differentiate through being naturally extendable (Foster, 2015a, b).

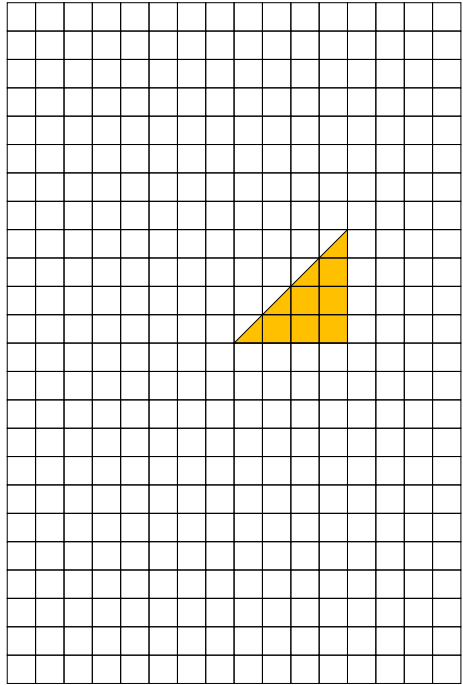
### 2.2.2 Devising equations etude

In this etude, students are asked to find integers  $a$ ,  $b$ ,  $c$  and  $d$  such that the equation  $ax + b = cx + d$  has integer solutions (Foster, 2012a), but this is presented to students in a more accessible way by using empty boxes rather than algebraic letters:  $\square x + \square = \square x + \square$ . In this task, the solution will be an integer if and only if  $(a - c)$  is a factor of  $(d - b)$  and  $a \neq c$ , but this level of generalisation is not necessarily expected. The intention is that students will experiment with different integers and discern some sense of what is possible, while at the same time gaining extensive practice at the technique of solving linear equations.

### 2.3 Enlargements etude

In this third etude, which addresses the topic of performing an enlargement of a given shape, students are presented with the diagram shown in Fig. 2, containing a right-angled isosceles triangle on a squared grid. The task is to find the locus of all possible positions for a centre of enlargement such that, for a scale factor of 3, the image produced lies completely on the grid. Students can generally find, without too much difficulty, one centre of enlargement that will work, but finding all possible points is demanding and may entail reverse reasoning from the possible image vertex positions to those of the original triangle. Further extensions are possible by considering different starting shapes, different positions of the starting shape on the grid and different scale factors. In all of this work, the enlargement procedure is being practised extensively within a wider investigative context.

**Fig. 2** Enlargement etude grid.  
(Taken from Foster, 2013d)



## 2.4 Summary

Each of the three etudes described above is intended to generate extensive opportunities for practising a single specified procedure within a rich problem-solving context. However, although etudes might be anticipated to have benefits for students in terms of greater engagement and creative problem solving, it is not known how effective they are in comparison with the standard approach of traditional exercises in the narrow objective of developing students' procedural fluency. It might be thought that incorporating other aspects beyond repetition of the desired procedure might to some extent diminish the effectiveness of a task for developing students' procedural fluency. However, the opposite could be the case if the problem-solving context to some extent directs students' attention away from the performance of the procedure and onto conceptual aspects, leading to greater automation. Consequently, the research question for these studies is: Are etudes as effective as traditional exercises at developing students' procedural fluency or not?

In these exploratory studies, it is important to emphasise that a choice was made to compare etudes and exercises only in very narrow terms of procedural fluency. While it is likely that etudes offer other, harder-to-measure benefits for students, such as providing opportunities for creative, open-ended, inquiry-based exploration and problem solving, unless they are at least about as good as traditional exercises at developing students' procedural fluency, it is unlikely, in a high-stakes assessment culture, that schools and teachers will feel able to use them regularly as an alternative. Traditional exercises are widely used by teachers not because they are perceived to be imaginative and creative sets of tasks but because they are believed to work in the narrow sense of developing fluency at necessary procedures. If there were some other way to achieve this, that did not entail the tedium of repetitive drill, it would presumably be

preferred—provided that it were equally effective at the main job. For this reason, in these studies the focus was entirely on the effect of etudes on procedural fluency. Rather than trying to measure the plausible but more nebulous ways in which etudes might be superior, in this first exploratory set of studies it was decided to focus solely on the question of the effectiveness of etudes for the purpose of developing procedural fluency.

### 3 Study 1: Expression polygons

The aim of this study was to investigate whether a particular etude (“Expression polygons”, see Section 2.2.1) is as effective as traditional exercises at developing students’ procedural fluency in solving linear equations, relative to the alternative hypothesis that the etude and the exercises are not equally effective.

#### 3.1 Method

A quasi-experimental design was used, with pairs of classes at the same school assigned to either the intervention (the etude) or control (traditional exercises) condition. Data was collected across one or two lessons, in which students in the intervention group tackled an etude while those in the control group worked through as many traditional exercises as possible in the same amount of time. Pre- and post-tests were administered at the beginning and end of the lesson(s).

A classical  $t$  test on the gain scores (post-test – pre-test) would be suitable for detecting a statistically significant difference between the two groups; however, within the paradigm of null hypothesis significance testing, failure to find such a difference would *not* constitute evidence *for* the null hypothesis of no difference—it would simply be inconclusive (Dienes, 2014). No evidence of a difference is not evidence of no difference. This is because failure to detect a difference might be a consequence of an underpowered study, which might have been able to detect a difference had a larger sample size (or more sensitive test) been used. For this reason, it is necessary to use a Bayesian approach for these studies, in order to establish how likely is the hypothesis of no difference between the two treatments (etude and traditional exercises) in terms of gain in procedural fluency, relative to the alternative hypothesis that there *is* a difference. Thus, Bayesian  $t$  tests were carried out on the gain scores obtained in each study, as described below.

##### 3.1.1 Instrument

The “Expression polygons” etude (Foster, 2012a, 2013a, 2015a) discussed in Section 2.2.1 was used for the intervention groups, and the control groups were provided with traditional exercises and asked to complete as many as possible in the same amount of time (see Fig. 3 for both), normally around 20 min. The exercises consisted of linear equations in which the unknown appears on both sides, leading to small integer solutions. Pre- and post-tests were designed (Fig. 4), each consisting of four equations of the same kind as those used in the exercises. In this way, it was hoped that any bias in the focus of the tests would be toward the control group (exercises), since the matching between the exercises and the post-test was intended to be as close as possible. Each test was scored out of 4, with one mark given for the correct solution to each equation. The post-test included a space at the end for open comments,

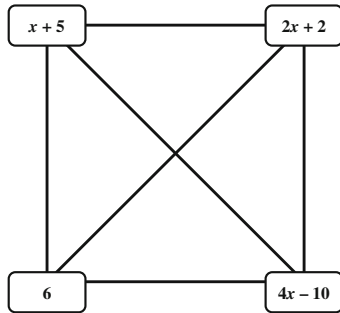


## Solving Equations

## Expression Polygons

In the diagram below, every line creates an equation.

So, for example, the line at the top gives the equation  $x + 5 = 2x + 2$ .



1. Write down and solve the six equations in this diagram.
2. What do you notice about your six solutions?
3. Now make up another diagram like this containing different expressions. Try to make the solutions to your *expression polygon* a "nice" set of numbers.
4. Make up some more *expression polygons* like this and see if other people can solve them.

Solve these equations.

Show your method for each one.

- |    |                       |    |                    |
|----|-----------------------|----|--------------------|
| 1  | $2x + 4 = 3x + 1$     | 16 | $7x - 3 = 2x + 2$  |
| 2  | $3x + 5 = 4x + 3$     | 17 | $3x - 5 = x + 1$   |
| 3  | $4x + 3 = 2x + 5$     | 18 | $x + 6 = 2x - 5$   |
| 4  | $2x - 3 = x - 1$      | 19 | $3x - 4 = x - 6$   |
| 5  | $2x + 1 = 3x - 2$     | 20 | $3x + 9 = x - 5$   |
| 6  | $5x - 3 = 2x + 12$    | 21 | $6x - 4 = x + 16$  |
| 7  | $4x + 9 = 8x - 31$    | 22 | $x - 7 = 7x - 25$  |
| 8  | $2x + 40 = 12x - 110$ | 23 | $x + 5 = 4x - 4$   |
| 9  | $3x + 4 = 5x - 8$     | 24 | $6x + 5 = 3x - 7$  |
| 10 | $2x - 8 = 3x - 16$    | 25 | $x + 1 = 7x - 17$  |
| 11 | $x + 1 = 5x + 9$      | 26 | $3x - 4 = 5x + 6$  |
| 12 | $5x = 2x + 12$        | 27 | $8x + 3 = 6x + 15$ |
| 13 | $9x + 8 = 20 - 3x$    | 28 | $x = 20 - x$       |
| 14 | $5x - 2 = x + 2$      | 29 | $3x - 1 = x + 7$   |
| 15 | $4x + 2 = 3x + 9$     | 30 | $x - 6 = 9 - 2x$   |

**Fig. 3** Study 1 materials: expression polygons etude (intervention) and traditional exercises (control)

asking students to write down “what you think about the work you have done on solving equations”. This question was intended to capture students’ perceptions of the two different tasks.

### 3.1.2 Participants

Schools were recruited through a *Twitter* request for help, and schools A, B and C (Table 1) took part in this study. These schools were a convenience sample, spanning a range of sizes and composition. In most schools in England, mathematics classes are set by attainment, and this was the case for schools A and B, while school C used mixed attainment classes. In all of the schools, teachers were asked to:

choose two similar classes (e.g. Year 8 or 9 parallel sets) who you are teaching to solve linear equations with the unknown on both sides (e.g. equations like  $7x - 1 = 5x + 3$ ). In these materials, all the solutions are whole numbers, but some may be negative.

A total of 241 mathematics students from Years 8 and 9 (age 12–14) participated. Forty-eight students’ pre- and post-tests could not be matched, either because they were not present for one of them or (for the vast majority) because they did not put their name clearly on the test. These students’ tests were excluded from the analysis, leaving  $N = 193$ . The large number of tests which could not be matched here was mainly due to the fact that in one particular class (20 students) none of the students wrote their names on either of their tests, and so none of the data from this class could be used.

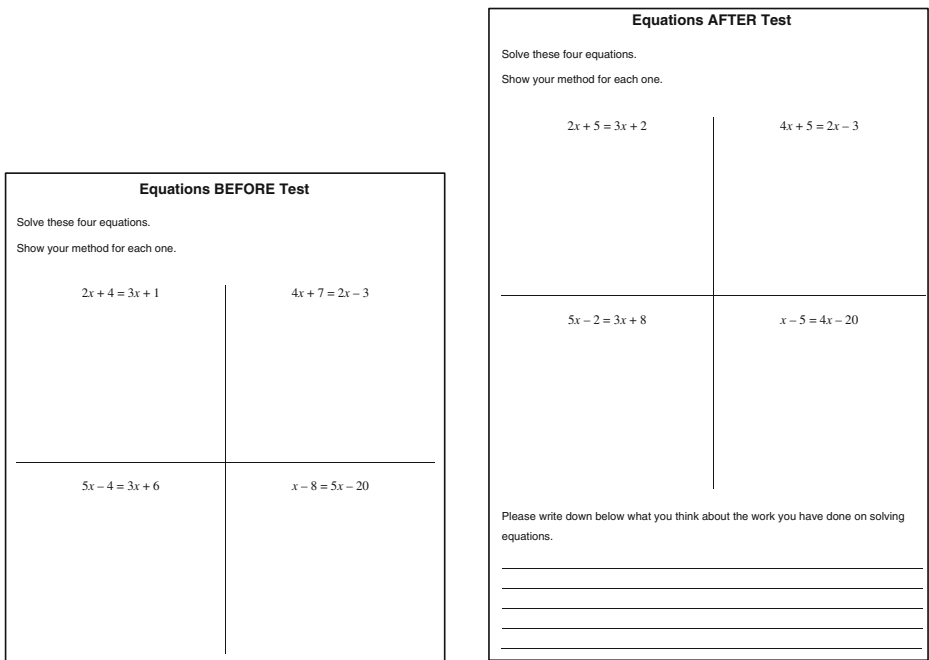


Fig. 4 Study 1 pre-test and post-test

### 3.1.3 Administration

Teachers were asked to use the materials with a pair of “parallel” classes across one or preferably two of the students’ normal mathematics lessons. Allocation was at class level, and schools were responsible for choosing pairs of classes that they regarded as similar, which were normally a pair from the same Year group which were setted classes at the same level (e.g., both set 3 out of 6). In most cases, the same teacher taught both classes, so as to minimise teacher effects.

Table 1 Data on participating schools and students (all three studies)

| Study | School | Location      | Type          | Sex   | Number of students |     |     | School total | Study total |
|-------|--------|---------------|---------------|-------|--------------------|-----|-----|--------------|-------------|
|       |        |               |               |       | Y8                 | Y9  | Y10 |              |             |
| 1     | A      | London        | Academy       | Mixed | 76                 |     |     | 76           | 193         |
|       | B      | West Midlands | Academy       | Mixed | 26                 | 25  |     | 51           |             |
|       | C      | West Midlands | Academy       | Girls | 20                 | 46  |     | 66           |             |
| 2     | D      | Scotland      | Comprehensive | Mixed | 29                 |     |     | 29           | 194         |
|       | E      | London        | Academy       | Mixed | 27                 |     |     | 27           |             |
|       | F      | East Midlands | Academy       | Mixed | 86                 |     |     | 86           |             |
|       | G      | East Midlands | Academy       | Mixed |                    | 18  |     | 18           |             |
| 3     | H      | Kent          | Academy       | Mixed |                    | 34  |     | 34           | 141         |
|       | I      | West Midlands | Academy       | Mixed |                    | 52  |     | 52           |             |
|       | J      | Oxfordshire   | Academy       | Mixed |                    |     | 47  | 47           |             |
|       | K      | West Midlands | Comprehensive | Mixed |                    |     | 42  | 42           |             |
|       |        |               |               |       |                    |     | 42  | 42           |             |
|       |        |               |               |       | 264                | 175 | 89  | 528          | 528         |

Y8, Y9 and Y10 are students aged 12–13, 13–14 and 14–15, respectively

Pre- and post-tests were administered individually in the same amount of time and until most students had finished (normally about 10 min. for each). Both classes were then taught by the teacher how to solve linear equations with the unknown on both sides. Teachers were asked to teach both classes “as you would normally, in the same way, and for approximately the same amount of time”. Following this, the control class received traditional exercises (Fig. 3), with the expectation that the number of questions would be more than enough for the time available (normally about 20 min.) and that students would not complete all of them, which generally proved to be the case. The intervention group received the “Expression polygons” etude (Fig. 3). Teachers were advised that “It is important that [the students] go beyond solving the six equations and spend some time generating their own expression polygons (or trying to).” Teachers were asked to allow the two classes the same amount of time to work on these tasks: “however much time you have available and feel is appropriate; ideally at least a whole lesson and perhaps more”. It is estimated that this was generally about 20–30 min. During this phase, teachers were asked to help both classes as they would normally, using their professional judgement as to what was appropriate, so that the students would benefit from the time that they spent on these tasks. Then the post-test was administered in the same way as the pre-test.

### 3.2 Results

The mean and standard deviation of the scores for both conditions at pre- and post-test, along with mean gain scores calculated as the mean of (post-test – pre-test) for each student, are shown in Table 2 and Fig. 5. The similarity of the mean scores on the pre-test is reassuring regarding the matching of the parallel classes. A Bayesian  $t$  test was carried out on the gain scores, using the *BayesFactor*<sup>2</sup> package in *R*, comparing the fit of the data under the null hypothesis (the etude is as effective as the traditional exercises) and the alternative hypothesis (the etude and the exercises are not equally effective). A Bayes factor  $B$  indicates the relative strength of evidence for two hypotheses (Dienes, 2014; Rouder, Speckman, Sun, Morey, & Iverson, 2009), and means that the data are  $B$  times as likely under the null hypothesis as under the alternative. With a Cauchy prior width of .707, an estimated Bayes factor (null/alternative) of 1.03 was obtained, indicating no reason to conclude in favour of either hypothesis. (Conventionally, a Bayes factor between 3 and 10 represents “substantial” evidence [Jeffreys, 1961].) Prior robustness graphs for all of the Bayesian analyses described in this paper are included in the Appendix. In this case, calculation indicates that an exceptionally wide Cauchy prior width of more than 2.39 would be needed in order to obtain a “substantial” (Jeffreys, 1961) Bayes factor. The 95% credible interval<sup>3</sup> for the standardised effect size was [–.545, .005].

Students’ comments on the study were few and generally related to the teaching episode rather than the etude or exercises. Insufficient responses meant that analysis of students’ perceptions of the two tasks was not possible.

### 3.3 Discussion

Such an inconclusive result does not allow us to say that either the exercises or the etude is superior in terms of developing procedural fluency, and neither does it allow us to say that there is evidence of no difference. Scrutiny of the students’ work suggested that in the time

<sup>2</sup> <http://bayesfactorpcl.r-forge.r-project.org/>

<sup>3</sup> A *credible interval* is the Bayesian analogue of a frequentist confidence interval. A credible interval has the much simpler interpretation that the probability of the parameter lying in the 95% credible interval is .95. The 95% highest-density intervals were also calculated for the standardised effect size for each study, and in each case were extremely close to the 95% credible intervals, and so are not reported.

**Table 2** Scores for study 1 (Expression polygons)

|           | $M_{pre}$ | $M_{post}$ | $M_{gain}$ | $N$ | $SD_{pre}$ | $SD_{post}$ | $SD_{gain}$ |
|-----------|-----------|------------|------------|-----|------------|-------------|-------------|
| Etude     | 1.45      | 1.99       | .542       | 107 | 1.38       | 1.33        | 1.34        |
| Exercises | 1.41      | 2.35       | .942       | 86  | 1.58       | 1.38        | 1.46        |

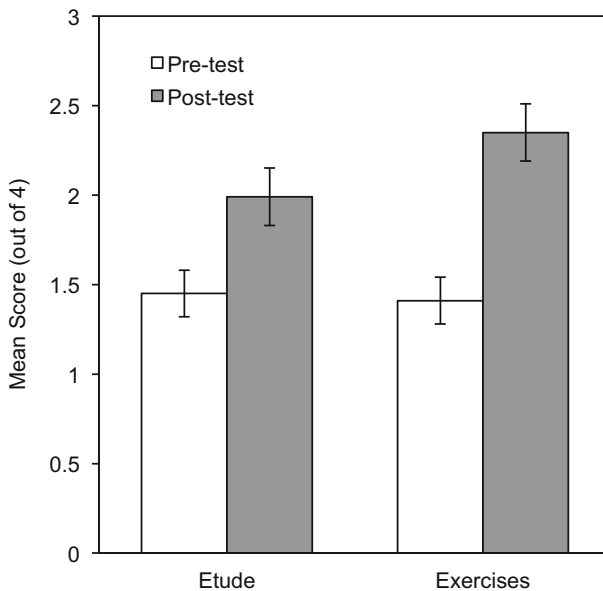
available many had engaged only superficially with the etude, whereas students in the control group had generally completed many exercises. It is possible that the style of the etude task was unfamiliar and/or that students were unclear regarding what they were supposed to do. For this reason, it was decided to devise a new etude to address the same topic of linear equations, one that it was hoped would be easier for students to understand and more similar in style to tasks that they might be familiar with. This etude formed the basis of study 2.

### 4 Study 2: Devising equations

The aim of this study was to investigate whether a different etude (“Devising equations”, see Section 2.2.2) is as effective as traditional exercises at developing students’ procedural fluency in solving linear equations, relative to the alternative hypothesis that the etude and the exercises are not equally effective.

#### 4.1 Method

The same quasi-experimental design was used as in study 1, with pairs of classes at the same school assigned to either the intervention (the “Devising equations” etude, see Section 2.2.2) or control (traditional exercises).



**Fig. 5** Study 1 results. (Error bars indicate ± 1 standard error)

### 4.1.1 Instrument and administration

This time the intervention group received the “Devising equations” etude, as described in Section 2.2.2 (see Fig. 6). The control group received the same set of traditional exercises as used in study 1 (see Fig. 3) and were asked to complete as many as possible in the same amount of time as given to the etudes group. The same pre- and post-tests were used as in study 1 (see Fig. 4). Administration was exactly as for study 1, except that this time the only advice given to teachers regarding the etude was that students should “generate and solve their own equations”.

### 4.1.2 Participants

Schools were again recruited through a *Twitter* request. Schools D, E, F, G and H (Table 1) took part, all of which used attainment setting for mathematics. Teachers were again asked to choose parallel classes, and a total of 213 mathematics students from Years 8 and 9 (age 12–14) participated. This time, 19 students’ pre- and post-tests could not be matched, because students did not always put their names on their tests, leaving  $N = 194$ .

## 4.2 Results

Results are shown in Table 3 and Fig. 7. As in study 1, a Bayesian  $t$  test was carried out on the gain scores (Dienes, 2014; Rouder et al., 2009), with a Cauchy prior width of .707, this time giving a Bayes factor (null/alternative) of 5.92. This means that the data are nearly six times as

### Equations Task

1. Make up an equation by choosing numbers to go in the empty boxes.

$$\square x + \square = \square x + \square$$

For example, if you chose the numbers **5**, **4**, **2** and **10**, you would get the equation  $5x + 4 = 2x + 10$ .

2. Solve your equation.

For example, when you solve the equation  $5x + 4 = 2x + 10$  you get  $x = 2$ .

3. Does your equation have a whole-number answer like this one?

4. Choose another set of our numbers to make another equation.

Try to make as many equations as you can that have whole-number answers.

**Fig. 6** Study 2 “Devising equations” task

**Table 3** Scores for study 2 (devising equations)

|           | $M_{pre}$ | $M_{post}$ | $M_{gain}$ | $N$ | $SD_{pre}$ | $SD_{post}$ | $SD_{gain}$ |
|-----------|-----------|------------|------------|-----|------------|-------------|-------------|
| Etude     | 1.98      | 2.63       | .647       | 75  | 1.55       | 1.48        | 1.21        |
| Exercises | 1.58      | 2.29       | .710       | 119 | 1.54       | 1.48        | 1.27        |

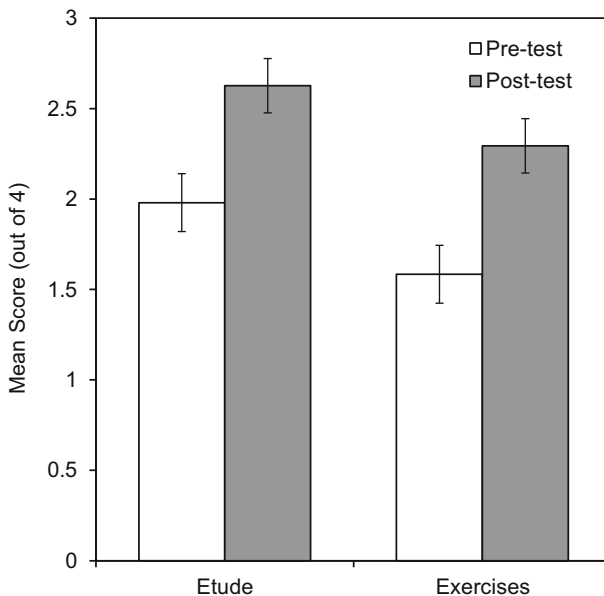
likely under the null hypothesis (the etude is as effective as the traditional exercises) as under the alternative hypothesis (the etude and the exercises are not equally effective). Conventionally, a Bayes factor between 3 and 10 represents “substantial” evidence (Jeffreys, 1961). The prior robustness graph (see Appendix) indicates that any Cauchy prior width of more than .317 would have led to a Bayes factor of at least 3, which suggests that this finding is robust. The 95% credible interval for the standardised effect size was [-.326, .233].

Again, students’ comments were insufficiently plentiful or focused on the task to enable an analysis.

### 4.3 Discussion

Study 2 provides substantial evidence that there is little difference across one or two lessons between the effect on students’ procedural fluency of using traditional exercises or the “Devising equations” etude. Examination of students’ work showed a much greater engagement with this etude than with the “Expression polygons” one used in study 1, as evidenced by far more written work, so it is plausible that the effect of this etude might consequently have been stronger and, in this case, was closely matched to that of the exercises.

In an attempt to extend the bounds of generalisability of this finding, a third study was conducted, using the enlargements etude discussed in Section 2.3, in order to see whether a similar result would be obtained in a different topic area.



**Fig. 7** Study 2 results. (Error bars indicate ± 1 standard error)

## 5 Study 3: Enlargements

The aim of this study was to investigate whether a third etude (“Enlargements”, see Section 2.3) is as effective as traditional exercises at developing students’ procedural fluency in a different (geometric) topic area: performing an enlargement of a given shape on a squared grid with a specified positive integer scale factor. As before, the alternative hypothesis was that the etude and the exercises are not equally effective.

### 5.1 Method

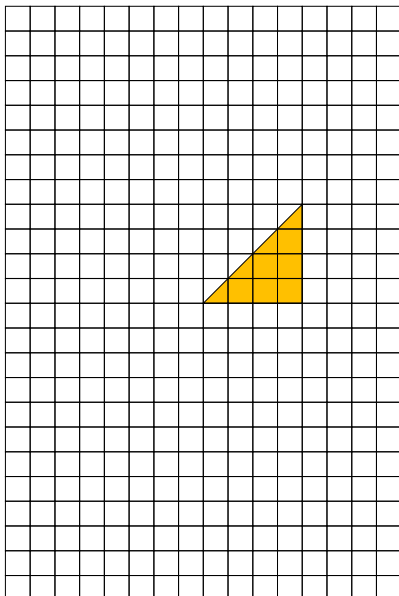
The same quasi-experimental design was used as in studies 1 and 2, with pairs of parallel classes in each school assigned to either the intervention (this time the “Enlargements” etude) or control (traditional exercises) condition.

#### 5.1.1 Instrument and administration

The “Enlargements” etude (Foster, 2013d) discussed in Section 2.3 was used for the intervention groups, and the control groups were provided with traditional exercises and asked to complete as many as possible in the same amount of time (see Fig. 8 for both). The exercises consisted of a squared grid containing five right-angled triangles and four given points. Each question asked students to enlarge one of the given shapes by a scale factor of 2, 3 or 5, using as centre of enlargement one of the given points. Pre- and post-tests were administered (Fig. 9), in which students were asked to enlarge a given triangle with a scale factor of 4 on a squared grid about a centre of enlargement marked with a dot.

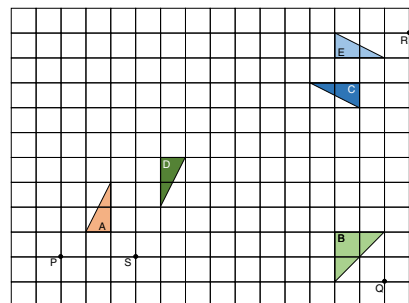
#### Enlargement Task

For a scale factor 3 enlargement of this triangle, where can the centre of enlargement be so that all of the enlarged shape is on the grid?



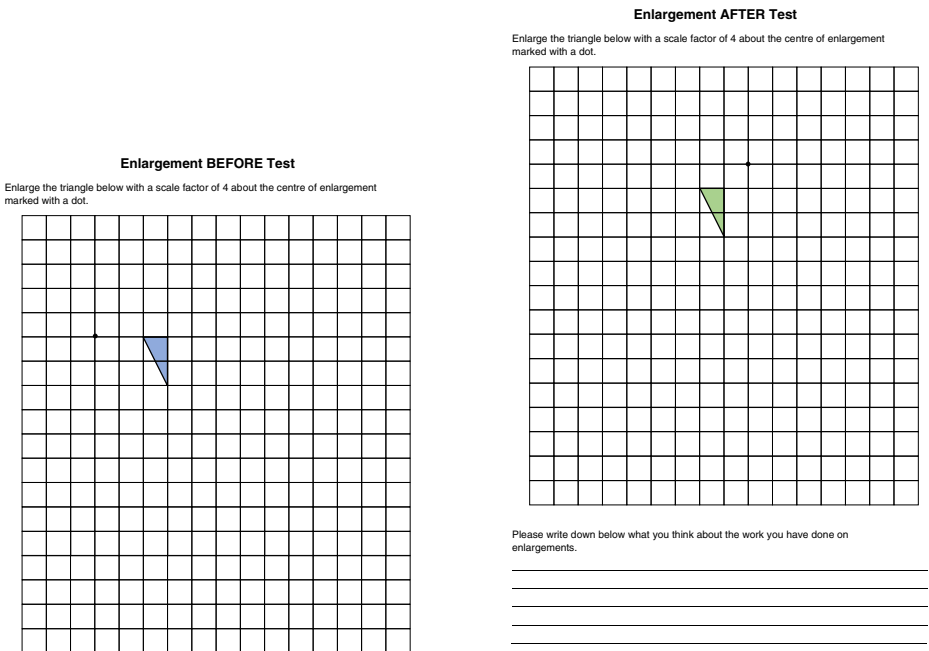
#### Enlargement Exercises

Here are five shapes (A, B, C, D and E) and four points (P, Q, R and S).



1. Enlarge shape A by a scale factor of 3 about centre of enlargement P. Label your shape F.
2. Enlarge shape B by a scale factor of 2 about centre of enlargement Q. Label your shape G.
3. Enlarge shape C by a scale factor of 3 about centre of enlargement R. Label your shape H.
4. Enlarge shape A by a scale factor of 2 about centre of enlargement S. Label your shape I.
5. Enlarge shape D by a scale factor of 2 about centre of enlargement S. Label your shape J.
6. Enlarge shape E by a scale factor of 5 about centre of enlargement R. Label your shape K.
7. Enlarge shape D by a scale factor of 2 about centre of enlargement P. Label your shape L.

**Fig. 8** Study 3 materials: enlargement etude (intervention) and traditional exercises (control)



**Fig. 9** Study 3 pre-test and post-test

The pre- and post-tests were intended to be as similar as possible in presentation to the traditional exercises, again in the hope that any bias in the focus of the post-test would be in favour of the control group. As before, the post-test included a space at the end for open comments, asking students to write down “what you think about the work you have done on enlargements”. Each test was scored out of 4, with one mark for each correctly positioned vertex and one for an enlarged triangle of the correct shape, size and orientation (not necessarily position). Administration was exactly as for studies 1 and 2, except that this time teachers were asked to ensure

that the students understand that they are meant to try to find *as many possible positions for the centre of enlargement as they can—perhaps even the whole region where these centres can be*. Students could also go on to explore what happens if the starting triangle is in a different position, or is a different shape, or if a different scale factor is used (original emphasis).

The purpose of this was to try to ensure that the students would engage extensively with the etude and not assume that finding one viable centre of enlargement was all that was required.

### 5.1.2 Participants

As before, schools were recruited through a *Twitter* request. Schools I, J and K in Table 1 took part, all of which used attainment setting for mathematics lessons. Teachers were again asked to choose parallel classes, and a total of 151 mathematics students from Years 9 and 10 (age 13–15) participated. Year 9–10 classes were used this time, rather than Year 8–9 classes, due to teachers’ choices about suitability for this different topic. This time, only 10 students’ pre- and post-tests could not be matched, again because of missing names on some of the tests, leaving  $N = 141$ .



**Table 4** Scores for study 3 (enlargement)

|           | $M_{pre}$ | $M_{post}$ | $M_{gain}$ | $N$ | $SD_{pre}$ | $SD_{post}$ | $SD_{gain}$ |
|-----------|-----------|------------|------------|-----|------------|-------------|-------------|
| Etude     | 2.05      | 3.45       | 1.39       | 74  | 1.63       | 1.24        | 1.67        |
| Exercices | 2.24      | 3.73       | 1.49       | 67  | 1.64       | .90         | 1.58        |

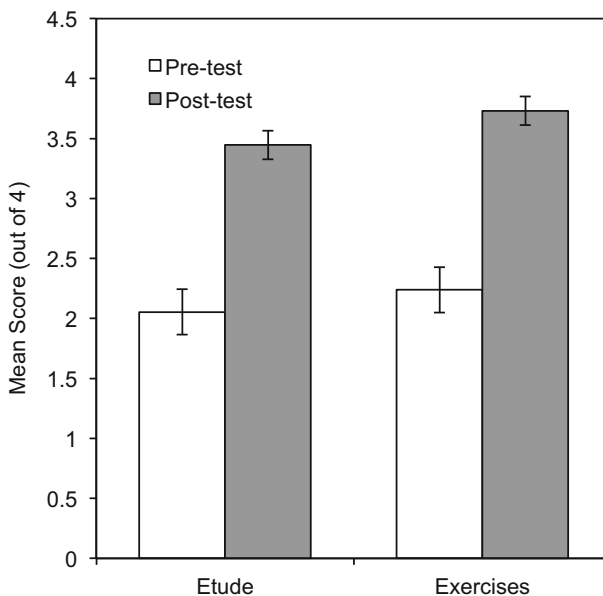
## 5.2 Results

Analysis proceeded as before, and the results are shown in Table 4 and Fig. 10. Again, a Bayesian  $t$  test was carried out on the gain scores (Dienes, 2014; Rouder et al., 2009), with a Cauchy prior width of .707, this time giving a Bayes factor (null/alternative) of 5.20, meaning that the data are about five times as likely under the null hypothesis (the etude is as effective as the traditional exercises) as under the alternative hypothesis (the etude and the exercises are not equally effective). Conventionally, a Bayes factor between 3 and 10 represents “substantial” evidence (Jeffreys, 1961). The prior robustness graph (see Appendix) indicates that any Cauchy prior width of more than .365 would have led to a Bayes factor of at least 3, which suggests that the finding is robust. The 95% credible interval for the standardised effect size was  $[-.384, .257]$ .

Once again, student comments were too few to allow a reasonable analysis.

## 5.3 Discussion

Study 3 provides substantial evidence that there is no difference across one or two lessons between the effect on students’ procedural fluency of using traditional exercises or this enlargement etude. Examination of students’ work showed a lot of drawing on the sheets, with many students correctly finding the locus of all possible positions for the centre of enlargement. It may be that this greater



**Fig. 10** Study 3 results. (Error bars indicate  $\pm 1$  standard error)

**Table 5** Bayes factors (null/alternative) for each study, and combined

| Study    | Bayes factor |
|----------|--------------|
| 1        | 1.03         |
| 2        | 5.92         |
| 3        | 5.20         |
| Combined | 5.83         |

degree of engagement (relative to study 1) could account for this etude being of comparable benefit to the exercises, as was the case in study 2.

## 6 General discussion

The Bayes factors obtained in these three studies were combined using the *BayesFactor* package in *R* and the *meta.ttestBF* Bayesian meta-analysis function. Again using a Cauchy prior width of .707, this time an estimated combined Bayes factor (null/alternative) of 5.83 was obtained (Table 5). This again falls within the conventionally accepted range of 3 to 10 for “substantial” evidence (Jeffreys, 1961). This means that, taken together, the three studies reported support the conclusion in favour of the null hypothesis that the etudes are as effective as the traditional exercises in developing students’ procedural fluency, relative to the alternative hypothesis that the etudes and the exercises are not equally effective.

The smaller Bayes factor for study 1 may have resulted from a less clearly articulated etude that was unfamiliar in style to the students, requiring a greater degree of initiative in constructing expressions than is normally expected in mathematics classrooms. If it is the case that students were less sure what was expected of them, this could explain why the etudes group carried out less equation solving here than the exercises group did. As reported above, in studies 2 and 3, a greater effort was made in the teacher instructions to explain the intentions of the task, and a greater engagement with the etudes was inferred from the quantity of written work produced.

## 7 Conclusion

These three exploratory studies suggest that the etudes trialled here are as effective as the traditional exercises in developing students’ procedural fluency. Consequently, for a hypothetical teacher whose sole objective was to develop students’ procedural fluency, it should be a matter of indifference whether to do this by means of exercises or etudes. Given the plausible benefits of etudes in terms of richness of experience and opportunity for open-ended problem solving and creative thinking, it may be that etudes might on balance be preferred (Foster, 2013b).

It should be stressed that only three etudes were tested in these studies across only two mathematics topics and with students aged 12–15. Further studies using other etudes in other topic areas and with students outside this age range would be necessary to extend the generalisability of this finding. In addition, studies including delayed post-tests would be highly desirable, but were not practicable for this initial exploratory study. It would also be important to examine evidence for the hypothesised *benefits* of etudes beyond the narrow focus of these studies on procedural fluency. For example, it is plausible that etudes are more engaging for students, provide opportunities for students to operate more autonomously and

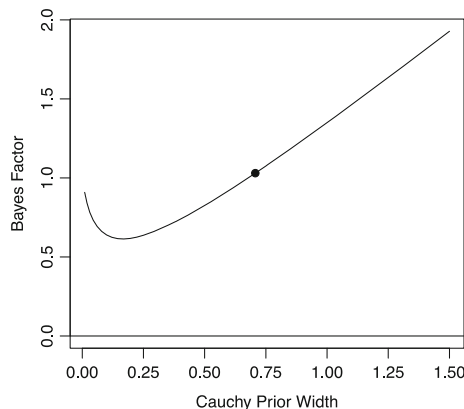
solve problems, promote discussion and reasoning and support conceptual understanding of the mathematics. Classroom observation data, other kinds of assessments, as well as canvassing teacher and student perspectives, would be necessary to explore the extent to which this might be the case.

Caution must be exercised in interpreting these findings, since the constraints of the participating schools did not allow random allocation of students to condition (etude or exercises). Instead, schools selected pairs of “parallel” classes, generally based on level of class within the Year group (e.g., set 3 out of 6). It is reassuring that pre-test scores were generally close across the two conditions, but there remains the possibility that the parallel classes differed on some relevant factor. It should also be noted that some pre- and post-tests could not be matched, as students did not write their names on their tests, meaning that these tests had to be excluded from the data. In studies 2 and 3, the percentages of tests excluded were 9% and 7%, respectively, but in study 1 the percentage was much higher (20%). However, this was largely the result of one particular class, in which none of the students wrote their names on either test; ignoring this class, the percentage of missing data was a less severe 12%. However, these higher than desirable percentages of missing data are a reason to be cautious in interpreting these findings.

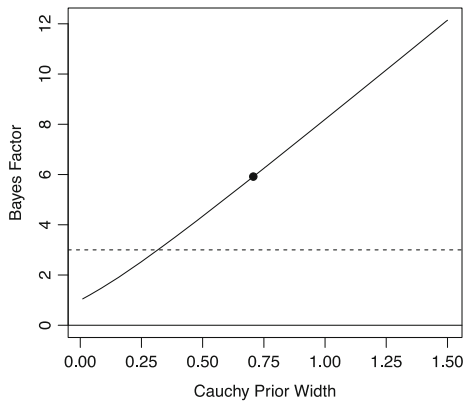
The extent of the guidance given to teachers about how to use the etudes was necessarily highly limited by the constraints of these studies. For practical reasons, the entire instructions on conducting the trials were restricted to one side of A4 paper. No professional development was involved, as these trials were carried out at a distance, and in most cases the participating schools and teachers were recruited via *Twitter* and contacted solely by email, and were not known personally to the researcher. It may be supposed that students would derive far greater benefit from etudes if they were deployed by teachers who had received professional development which involved prior opportunities to think about and discuss ways of working with these sorts of tasks. It remains for future work to explore this possibility.

## Appendix: Prior robustness graphs for the three Bayesian analyses

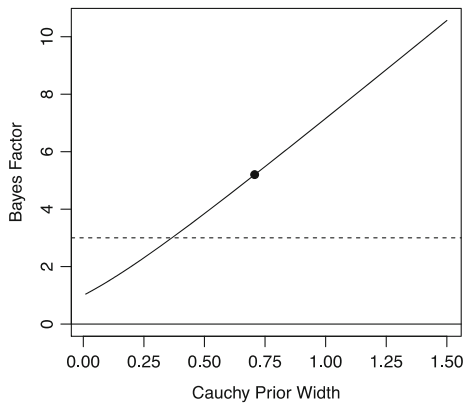
### Study 1.



Study 2.



Study 3.



Horizontal dashed lines show the conventional cut-off Bayes Factor of 3 for “substantial” evidence (Jeffreys, 1961).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

ACME. (2012). *Raising the bar: Developing able young mathematicians*. London: Advisory Committee on Mathematics Education. Retrieved from <http://www.acme-uk.org/media/10498/raisingthebar.pdf>. Accessed 23 Sep 2017.

Andrews, P. (2002). Angle measurement: An opportunity for equity. *Mathematics in School*, 31(5), 16–18.

Askew, M., Hodgen, J., Hossain, S., & Bretscher, N. (2010). *Values and variables: Mathematics education in high-performing countries*. London: Nuffield Foundation.

- Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning*. London: Routledge.
- Boaler, J. (2015). *Fluency without fear: Research evidence on the best ways to learn math facts*. Stanford, CA: youcubed. Retrieved from <https://www.youcubed.org/fluency-without-fear/>. Accessed 23 Sep 2017.
- Codding, R. S., Burns, M. K., & Lukito, G. (2011). Meta-analysis of mathematic basic-fact fluency interventions: A component analysis. *Learning Disabilities Research & Practice, 26*(1), 36–47.
- Department for Education (DfE). (2013). *Mathematics programmes of study: Key stage 3: National curriculum in England*. London: DfE.
- Dienes, Z. (2014). Using Bayes to get the most out of non-significant results. *Frontiers in Psychology, 5*, 781.
- Encyclopaedia Britannica. (2007). *Étude*. Retrieved from <https://www.britannica.com/art/etude-music>. Accessed 11 Feb 2017.
- Ericsson, A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. New York: Houghton Mifflin Harcourt.
- Fan, L., & Bokhove, C. (2014). Rethinking the role of algorithms in school mathematics: A conceptual model with focus on cognitive development. *ZDM Mathematics Education, 46*(3), 481–492.
- Foster, C. (2011). A picture is worth a thousand exercises. *Mathematics Teaching, 224*, 10–11.
- Foster, C. (2012a). Connected expressions. *Mathematics in School, 41*(5), 32–33.
- Foster, C. (2012b). Working without a safety net. *The Australian Mathematics Teacher, 68*(2), 25–29.
- Foster, C. (2013a). Connected quadratics. *Teach Secondary, 2*(1), 46–48.
- Foster, C. (2013b). Mathematical études: Embedding opportunities for developing procedural fluency within rich mathematical contexts. *International Journal of Mathematical Education in Science and Technology, 44*(5), 765–774.
- Foster, C. (2013c). Resisting reductionism in mathematics pedagogy. *Curriculum Journal, 24*(4), 563–585.
- Foster, C. (2013d). Staying on the page. *Teach Secondary, 3*(1), 57–59.
- Foster, C. (2014). Mathematical fluency without drill and practice. *Mathematics Teaching, 240*, 5–7.
- Foster, C. (2015a). Expression polygons. *Mathematics Teacher, 109*(1), 62–65.
- Foster, C. (2015b). The convergent–divergent model: An opportunity for teacher–learner development through principled task design. *Educational Designer, 2*(8).
- Foster, C. (2016). Confidence and competence with mathematical procedures. *Educational Studies in Mathematics, 91*(2), 271–288.
- Hewitt, D. (1996). Mathematical fluency: The nature of practice and the role of subordination. *For the Learning of Mathematics, 16*(2), 28–35.
- Hewitt, D. (2015). The economic use of time and effort in the teaching and learning of mathematics. In S. Oesterle & D. Allan (Eds.), *Proceedings of the 2014 Annual Meeting of the Canadian Mathematics Education Study Group* (pp. 3–23). Edmonton, Canada: University of Alberta.
- Hodgen, J. (2015). Unlocking potential in our curriculum: Mastering subjects. In A. Bingham (Ed.), *Lessons learned: Putting experience to work* (pp. 35–44). London: Ark, King's College London.
- Jeffreys, H. (1961). *The theory of probability* (3rd ed.). Oxford: Oxford University Press.
- Kling, G., & Bay-Williams, J. M. (2015). Three steps to mastering multiplication facts. *Teaching Children Mathematics, 21*(9), 548–559.
- Lemov, D., Woolway, E., & Yezzi, K. (2012). *Practice perfect: 42 rules for getting better at getting better*. San Francisco, CA: Jossey-Bass.
- Leung, F. K. (2014). What can and should we learn from international studies of mathematics achievement? *Mathematics Education Research Journal, 26*(3), 579–605.
- Li, S. (1999). Does practice make perfect? *For the Learning of Mathematics, 19*(3), 33–35.
- National Association of Mathematics Advisors [NAMA]. (2016). Five myths of mastery in mathematics. *Mathematics Teaching, 251*, 20–24.
- National Centre for Excellence in the Teaching of Mathematics [NCETM] (2016). *The essence of maths teaching for mastery*. Retrieved from <https://www.ncetm.org.uk/files/37086535/The+Essence+of+Maths+Teaching+for+Mastery+june+2016.pdf>. Accessed 23 Sep 2017.
- National Council of Teachers of Mathematics [NCTM]. (2014). *Procedural fluency in mathematics: A position of the National Council of Teachers of Mathematics*. Reston, VA: NCTM.
- Office for Standards in Education [Ofsted]. (2012). *Mathematics: Made to measure*. London: Ofsted. Retrieved from [https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/417446/Mathematics\\_made\\_to\\_measure.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/417446/Mathematics_made_to_measure.pdf). Accessed 23 Sep 2017.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237.
- Taleporos, E. (2005). Consequential validity: A practitioner's perspective. *Educational Measurement: Issues and Practice, 17*(2), 20–23.

- Truss, E. (2013, January 17). *Education minister Elizabeth Truss speaks at the North of England Education Conference*. Retrieved from <https://www.gov.uk/government/speeches/education-minister-elizabeth-truss-speaks-at-the-north-of-england-education-conference> . Accessed 11 Feb 2017.
- Watson, A., & De Geest, E. (2014). Department-initiated change. *Educational Studies in Mathematics*, 87(3), 351–368.
- Willingham, D. T. (2009). *Why don't students like school: A cognitive scientist answers questions about how the mind works and what it means for the classroom*. San Francisco, CA: Jossey-Bass.
- Yeo, J. B. (2017). Development of a framework to characterise the openness of mathematical tasks. *International Journal of Science and Mathematics Education*, 15(1), 175–191.