Learning and Individual Differences 37 (2015) 107–117

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Third graders' verbal reports of multiplication strategy use: How valid are they?

Helen C. Reed [a,*], Claire Stevenson [a,1], Marije Broens-Paffen [a,2], Paul A. Kirschner [b], Jelle Jolles [a]

[a] Department of Educational Neuroscience and LEARN! research institute for learning and education, Faculty of Psychology and Education, VU University Amsterdam, Van der Boechorststraat 1, 1081 BT Amsterdam, The Netherlands

[b] Welten Institute, Research Centre for Learning, Teaching and Technology, Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands

ARTICLE INFO

Article history:
Received 8 January 2014
Received in revised form 20 September 2014
Accepted 7 November 2014
Available online xxxx

Keywords:
Verbal reports
Multiplication strategies
Multiplication performance
Problem characteristics
Cognitive characteristics

ABSTRACT

This study investigates whether children's verbal reports accurately represent their thinking processes when solving simple multiplication problems. A total of 106 third graders in Dutch mainstream primary schools solved simple multiplication problems and retrospectively reported how they had done this. The degree to which verbal reports predict children's problem-solving performance in ways that correspond to known patterns of response latency, accuracy, errors and strategy choice was assessed. The analyses took account of relevant problem characteristics and child cognitive characteristics (i.e., math ability, verbal ability, phonological decoding speed) known to affect the relation between strategy use and multiplication performance. The verbal reports were largely consistent with known patterns, supporting the use of verbal reports in assessing multiplication strategy use. Moreover, verbal reports provide valuable information that can alert teachers and educational researchers to specific issues that students face when solving simple multiplication problems. Considerations for soliciting reliable verbal reports are suggested.

1. Introduction

It is well established that children use a variety of strategies to solve arithmetic problems. Common strategies involve *counting*, *repeated addition* (e.g., $4 \times 3 = 3 + 3 + 3 + 3 = 12$), using *derived facts* (also known as decomposition or transformation, e.g., $3 + 4 = 3 + 3 + 1 = 7$; $3 \times 6 = 3 \times 5 + 3$) and *direct retrieval* of facts from memory (Imbo & Vandierendonck, 2008a,b; Lemaire & Siegler, 1995; Mabbott & Bisanz, 2003; Sherin & Fuson, 2005; Siegler, 1987, 1988; Siegler & Shipley, 1995).

Children show great variability in thinking and may at any one time use different strategies in different circumstances, depending on their age, problem difficulty, their degree of experience with the type of problem, their degree of confidence in the solution, strategy characteristics, and individual differences such as gender, achievement level and working memory capacity (Foxman & Beishuizen, 2002; Imbo & Vandierendonck, 2008a,b; Lemaire & Siegler, 1995; Mabbott & Bisanz, 2003; Siegler, 1988; Siegler & Shipley, 1995; Timmermans, Van Lieshout, & Verhoeven, 2007; Verschaffel, Luwel, Torbeyns, & Van Dooren, 2009). Typically developing children progress from relying on procedural strategies such as counting and addition to increasingly frequent use of more mature memory-based strategies, particularly direct retrieval (Geary, 2004; Lemaire & Siegler, 1995). Low math performers, however, exhibit developmental delay in their patterns of strategy use and may have long-lasting difficulties in using memory-based retrieval strategies (Geary, 2004; Geary, Hoard, Nugent, & Bailey, 2012; Jordan, Hanich, & Kaplan, 2003). Reciprocally, arithmetic performance appears to depend on strategy use, with increased use of the fastest and most accurate strategy (i.e., direct retrieval) producing faster and more accurate performance (Geary, 2004; Lemaire & Siegler, 1995).

Clearly, children should be helped to progress from using time-consuming and error-sensitive procedural methods to using more mature retrieval strategies. Such progression does not necessarily occur – even with typically developing children – when these strategies are not given explicit attention in school (Steel & Funnell, 2001). This issue is particularly relevant for low math performers, for whom progression is often delayed (Geary et al., 2012; Jordan et al., 2003). The performance disadvantage of these children may be compounded when immature strategy use appropriates cognitive resources – most importantly working memory – resulting in a reduced capacity to process higher-level aspects of mathematical learning (Raghubar, Barnes, & Hecht, 2010).

An important issue for educational practice is to determine which problem-solving strategies a child is currently using. If children who

* Corresponding author. Tel.: +31 20 598 7023; fax: +31 20 598 8902.
E-mail addresses: HC.Reed@vu.nl (H.C. Reed), cstevenson@fsw.leidenuniv.nl (C. Stevenson), Marije.Broens@ou.nl (M. Broens-Paffen), Paul.Kirschner@ou.nl (P.A. Kirschner), J.Jolles@vu.nl (J. Jolles).

[1] Present address: Faculty of Social and Behavioural Sciences, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands.

[2] Present address: Faculty of Psychology and Educational Sciences, Open University of the Netherlands, P.O. Box 2960, 6401 DL Heerlen, The Netherlands.

http://dx.doi.org/10.1016/j.lindif.2014.11.010
1041-6080/

consistently use more immature strategies than their peers can be readily identified in the classroom, teachers could take remedial action before more complex learning is compromised (Dowker, 2004; Gersten, Jordan, & Flojo, 2005). Importantly, teachers would then not have to wait for signs of failure to become established before referring children for specialised testing. A practicable and direct way for teachers to assess strategy use would be to ask children how they solve arithmetic problems. Yet, if this method is to be advocated, it is first necessary to establish whether what children report about how they have solved a problem accurately represents their thinking processes (i.e., veridicality). To that end, the goal of the present study is to investigate the veridicality of children's verbal reports of strategy use during simple multiplication problem-solving. Difficulties in multiplication learning are often reported by students and teachers (Kilpatrick, Swafford, & Findell, 2001; Steel & Funnell, 2001; Wallace & Gurganus, 2005), which makes it particularly relevant to investigate strategy use in this area.

1.1. Identifying strategy use through verbal reports

Verbal reports can be extremely valuable and may provide the most informative picture of cognitive processing in problem-solving (Fox, Ericsson, & Best, 2011; Robinson, 2001; Taylor & Dionne, 2000; Sherin & Fuson, 2005; Siegler, 1987, 1989). In one of the few extant studies to explicitly consider the validity of children's verbal reports of arithmetical processing, Robinson (2001) argued that verbal reports can provide unique and rich information on children's problem-solving strategies that can help to build and test more complete theories of cognitive development. Verbal reports can also help identify difficulties encountered during problem-solving and help determine ways to alleviate these difficulties. Moreover, verbal reports are highly suitable for investigating individual differences in arithmetic strategy use (e.g., Imbo & Vandierendonck, 2008a,b; Mabbott & Bisanz, 2003; Van der Ven, Boom, Kroesbergen, & Leseman, 2012). In this respect, verbal reports can be superior to other commonly used measures, namely latency and accuracy data. Several authors have demonstrated that using these data to infer strategy use can severely distort the picture of performance. For example, when latencies are aggregated across subjects, individual differences are obscured, and when latencies are aggregated across strategies, variability in strategy use and performance is masked (Cooney, Swanson, & Ladd, 1988; Siegler, 1987, 1989).

Verbal reports can provide accurate indications of mental processing under certain conditions. For assessing children's arithmetic problem-solving, a number of conditions should be met. First, the act of reporting should not change performance (so-called 'reactivity'). For example, when people 'think aloud' while performing a task, the drain on cognitive resources required for verbalisation of task-related processes can impair performance (Chin & Schooler, 2008; Robinson, 2001; Russo, Johnson, & Stephens, 1989). Second, as the time taken to execute the task is of interest, additional time-consuming processing (e.g., verbalisation) should not be undertaken concurrently (Ericsson & Simon, 1993; Russo et al., 1989). Third, when a task is carried out through automatic processing (e.g., direct retrieval), people may be unable to report their mental actions at the time (Kirk & Ashcraft, 2001; Taylor & Dionne, 2000) but may be able to do so retrospectively (Robinson, 2001). Taken together, it is likely that these conditions may best be approximated by soliciting verbal reports immediately after task completion. Retrospective reports are widely used in psychological research and appear to have good validity when tasks are of short duration and relevant task-specific processing traces are still available in short-term memory (Crutcher, 1994; Ericsson & Simon, 1993; Taylor & Dionne, 2000). Interestingly, there is neurophysiological evidence for the validity of retrospective reports in mental arithmetic with adults (Grabner et al., 2009; Grabner & De Smedt, 2011[3]).

[3] Although the authors call these reports concurrent, the procedure description makes clear that reports were obtained immediately after solving the problem.

Nonetheless, there are reasons to question whether verbal reports are a valid reflection of cognitive processing, particularly when children are the respondents. Verbal report depends on the ability to recognise and articulate thought processes, and individuals – especially children – vary widely in the degree to which they are able to do this. For example, Siegler and Stern (1998) found that 90% of second graders in their study were able to use a shortcut strategy on inversion problems some time before they were able to explicitly report using it. Alibali (1999) and Broaders, Cook, Mitchell, and Goldin-Meadow (2007) reported that many third and fourth graders demonstrate an understanding of strategies for solving mathematical equivalence problems that they are unable to verbalise. Even adults are often able to perform tasks without being able to articulate how they have done this, a phenomenon that is well known in sequence and second language learning (e.g., Neil & Higham, 2012; Sanchez, Gobel, & Reber, 2010; Williams, 2005).

Furthermore, retrospective reports in particular could be open to bias and fabrication (Ericsson & Simon, 1993; Russo et al., 1989; Taylor & Dionne, 2000). In this respect, children's retrospective reports of arithmetic strategy use may be inaccurate in several ways. Children may over-report strategies whose salience is high (Kirk & Ashcraft, 2001) due to a certain type of instruction given during interview or through emphasis being put on particular strategies in the classroom (Taylor & Dionne, 2000). In that case, children may believe that reporting how they solved a problem is a test of – and therefore should reproduce – what they are supposed to have learned. Also, children may be aware that they are not supposed to use more primitive strategies such as counting. They may therefore deliberately under-report use of these strategies in order to conform to what they believe to be socially desirable. Furthermore, if a solution strategy involves both faster (e.g., direct retrieval) and slower (e.g., calculation) processes, children may only report the slower and perhaps more easily verbalised process (Ericsson & Simon, 1993; Kirk & Ashcraft, 2001).

In summary, although retrospective reporting is a promising method for assessing children's arithmetic strategy use in the classroom from both a theoretical and a practical point of view, important questions remain to be answered as to its validity. Validity is called into question when some thoughts are not reported (i.e., errors of omission) or when thoughts that did not occur are reported (i.e., errors of commission) (Russo et al., 1989). To date, the evidence of validity when used with children is not conclusive: discrepancies have been found between children's retrospective reports and other measures using observational or chronometric data (e.g., Cooney & Ladd, 1992; Cooney et al., 1988; Siegler, 1987, 1989; but see Wu et al., 2008). Furthermore, retrospective reports may be biased or fabricated: children may be tempted – as an effect of salience or social desirability – to under-report or over-report the use of certain strategies. It is important to resolve this issue if retrospective reporting is to be used as an assessment method in the classroom. If children over-report the use of mature strategies and under-report the use of immature strategies, teachers will not be able to determine their actual level of expertise and identify those who really do lag behind their peers.

1.2. Strategy use and multiplication performance

Previous research investigating children's multiplication performance from the point of view of strategy use has produced several robust findings. Performance is reported to be faster and more accurate when children use retrieval compared to procedural strategies and, of the procedural strategies, use of derived facts is faster and more accurate than counting and addition (Lemaire & Siegler, 1995; Siegler, 1988; Siegler & Shipley, 1995; Steel & Funnell, 2001). Regarding types of errors made, these tend to be primarily multiplicand-related (i.e., multiples of one of the multiplicands, e.g., $8 \times 4 = 28$). To a lesser extent, 'close misses' (i.e., small errors within a distance of 10% from the correct answer, e.g., $8 \times 4 = 33$) occur on procedural strategies and tables-related errors (i.e., from a multiplication table other than those of the multiplicands, e.g., 8 × 4 = 35) on retrieval (Lemaire & Siegler, 1995; Siegler, 1987, 1988; Verguts & Fias, 2005).

When investigating the relation between strategy use and performance, it is important to take account of both problem characteristics and child cognitive characteristics that are known to influence this relation. Relevant problem characteristics are the so-called problem-size effect, the tie effect and the five effect (De Brauwer, Verguts, & Fias, 2006; Imbo & Vandierendonck, 2008a,b; LeFevre et al., 1996; Mabbott & Bisanz, 2003; Siegler, 1988; Verguts & Fias, 2005). The problem-size effect refers to the fact that solutions are longer and less accurate when problems involve larger multiplicands (e.g., 7 × 9) than smaller ones (e.g., 2 × 3). The tie effect is that problems with two identical multiplicands ('ties'; e.g., 6 × 6) are solved faster and more accurately than problems with unequal multiplicands. The five effect is that problems including 5 as a multiplicand ('fives') are solved faster and more accurately than other problems of comparable size. Moreover, the problem-size effect is weaker for ties and fives.

Problem characteristics also affect strategy use. Thus, retrieval is used more frequently on smaller problems while procedural strategies are used on larger problems (Imbo & Vandierendonck, 2008a,b; Lemaire & Siegler, 1995; Siegler, 1988; Steel & Funnell, 2001). Furthermore, ties and smaller fives are most often solved by retrieval, while larger fives are most often solved by repeated addition or use of table strings (i.e., a memorised string such as '5, 10, 15, 20') (LeFevre et al., 1996).

Children's cognitive characteristics can also affect both strategy use and multiplication performance. Children with higher math ability produce faster, more accurate performance, with more frequent use of retrieval rather than procedural strategies (Geary et al., 2012; Jordan et al., 2003). Speed of phonological decoding, which taps the ability to quickly retrieve phonological representations from long-term memory, is also a significant predictor of arithmetic performance (Dehaene, 2011; Fuchs et al., 2006; Hecht, Torgesen, Wagner, & Rashotte, 2001; Simmons & Singleton, 2008). This finding can be understood from neuroimaging studies showing that mental arithmetic and retrieval of arithmetic facts engage brain regions involved in phonological processing and verbal memory (Dehaene, 2011; Geary, 2004). Consequently, arithmetic facts are thought to be stored as phonological codes in long-term verbal memory and higher decoding skill should produce faster and more accurate arithmetic performance.

Phonological decoding is also highly correlated with other variables that are linked to arithmetic performance: so-called 'rate of access' to phonological information in long-term memory, and the phonological loop in working memory. In this view, procedural strategies involve accessing and retrieving phonological codes for numeric terms and operators from long-term memory and then maintaining and manipulating these codes in phonological working memory (Dehaene, 2011; De Smedt, Taylor, Archibald, & Ansari, 2010; Geary, 2004; Hecht et al., 2001; Imbo & Vandierendonck, 2007; Simmons & Singleton, 2008). For example, solving the problem '3 × 4 =' through repeated addition would involve retrieving the phonological codes for each number in the sequence 4, 8 and 12 and for the symbols '+' and '=' and representing the calculation in phonological working memory as 'four plus four is eight plus four is twelve'. Efficient access to these codes as well as efficient processing in working memory should enable this procedure to be carried out quickly and accurately; when either is poor, however, representations are likely to decay before the procedure is completed and performance will suffer (Geary, 2004; Hecht et al., 2001).

Finally, verbal ability is also linked to arithmetic skills (Durand, Hulme, Larkin, & Snowling, 2005; LeFevre et al., 2010; Vukovic & Lesaux, 2013) and could, moreover, affect the quality of what children are able to express through verbal reports. One way in which verbal abilities can affect arithmetic performance relates to the view that children come to use their verbal abilities in the form of inner speech as a tool for guiding, planning and regulating their thinking and behaviour (Vygotsky, 1934/1986; Winsler, Fernyhough, & Montero, 2009). Inner speech plays an important role when using counting and addition strategies in arithmetical calculation (Ostad, 2013; Ostad & Sorensen, 2007) and lower or impaired verbal skills are related to inner speech deficits (Lidstone, Fernyhough, Meins, & Whitehouse, 2009; Lidstone, Meins, & Fernyhough, 2012; Williams & Jarrold, 2010). Thus, children with lower verbal ability are likely to be less proficient than children with higher verbal ability when using counting and addition strategies.

1.3. The present study

The present study focuses on the domain of simple multiplication (i.e., single-digit multiplication and multiples of 10). While attention has been given to children's verbal reports of strategy use for addition and subtraction (e.g., Robinson, 2001; Siegler, 1987, 1989), few studies have explicitly considered the validity of children's verbal reports of multiplication strategy use. Those that have, have included only small numbers of participants (see Cooney & Ladd, 1992; Cooney et al., 1988; Dubé & Robinson, 2010). The results of those studies are equivocal, with some findings indicating that children's reports of multiplication strategy use are veridical and others suggesting the opposite.

The present study aims to augment the small body of research in this area. The study focuses on third grade children at an intermediate stage of multiplication learning when a wide range and mix of strategies are employed, so that there is sufficient performance variability to allow individual differences to be detected. Importantly, correlated responses within individuals are taken account of through multilevel techniques, thereby addressing the earlier described objections to previous research when data is aggregated across subjects or averaged across strategies.

The study investigates the extent to which verbal reports accurately represent third graders' thinking processes (i.e., veridicality) during simple multiplication problem-solving. Following the reasoning of previous research (e.g., Robinson, 2001), veridicality would be demonstrated if verbal reports of strategy use correspond to known performance patterns for those strategies. Thus, the study investigates the extent to which children's problem-solving performance (in terms of response latency, accuracy, and error type) is predicted by their verbal reports, taking account of problem characteristics and cognitive characteristics known to affect the relation between strategy use and performance. In addition, it is examined whether reported strategies correspond to known patterns of strategy choice for problems with different characteristics. On the basis of the literature discussed, the following patterns are expected:

- (P1) *Strategies and performance*: Response latencies should be shorter and accuracy higher for retrieval compared with procedural strategies; comparing procedural strategies, use of derived facts should produce shorter response latencies and higher accuracy than counting and addition strategies.
- (P2) *Problem characteristics*: Response latencies should increase and accuracy should decrease as problem size increases; this effect should be mitigated for problems involving ties or fives.
- (P3) *Cognitive characteristics*: Math ability, speed of phonological decoding and verbal ability should positively predict performance; performance on procedural strategies should be faster and more accurate when phonological decoding speed is higher; performance on counting and addition strategies should be faster and more accurate when verbal ability is higher.
- (P4) *Error types*: Both retrieval and procedural strategies should produce primarily multiplicand-related errors, with a smaller likelihood of 'close misses' on procedural strategies and tables-related errors on retrieval.

(P5) *Strategy choice versus problem characteristics*: Retrieval should be used on smaller problems and procedural strategies should be used on larger problems unless these involve ties or fives; ties and smaller fives should be solved by retrieval; larger fives should be solved by repeated addition or table strings.

2. Methods

2.1. Participants

Participants were 107 Grade 3 children from 15 classes of seven mainstream primary schools in the Netherlands (mean age = 8;11 years; *SD* = 6.15 months; range = 7;9 − 10;2 years; 43.9% boys). One participant was unresponsive and was consequently excluded from the study, leaving a final sample of 106 children. At the time of study (i.e., end of first semester in Grade 3), children had been exposed to multiplication for about a year, were familiar with repeated addition and had been introduced to diverse 'shortcut' strategies (i.e., derived facts or decomposition), thereby increasing the range of problem-solving approaches that they could employ.

Given the focus of the study on identifying children who lag behind their age-level peers, participants were primarily low to average math performers. A group of above-average to high performers was included for comparison. Children were selected for participation on the basis of their math ability level on a standardised, norm-referenced mathematics test developed by the Dutch Central Institute for Test Development (CITO). According to these norms, 20% of individuals of the age-referenced population are average performers, 20% are below average, and 20% are low performers (the remaining 40% is equally split among above average and high performers). Parents of selected children were asked for written permission for their child's participation in the study after being fully informed of the research goals and procedures. Descriptive statistics are given in Table 1.

2.2. Procedure

Participating children were interviewed by an experienced teacher who was trained in the interview protocol and materials. Children were interviewed individually in a separate room at their own school; interviews lasted 20 minutes on average. Children were told that the interviewer wanted to know how they solved multiplication problems. No further information was given about the research goals. A full verbatim protocol was used for the interviews and all interviews were digitally recorded.

Each child was first asked to solve four simple verbally-presented baseline problems (e.g., 2 × 2) to ensure that (s)he possessed at least minimal multiplication skill. Then, to ensure that the child understood what was required in terms of verbal report, (s)he was presented with two practice problems on cards. For each problem, the child was asked to first give the answer verbally and then write the answer on an individual worksheet. Directly afterwards, the interviewer asked, "*How did you solve this problem*?" No attention was drawn to any particular solution strategy, to avoid influencing participants' reports (Kirk & Ashcraft, 2001; Taylor & Dionne, 2000). However, if a child did not give any response or said "*I don't know*", the interviewer stimulated him/her to answer through the use of open questions as: "*What did you think when you solved the problem*?" or "*What happened in your head*?" Note that the practice problems were expected to invoke experiences of both direct retrieval (i.e., 1 × 10) and a procedural strategy (i.e., 4 × 6) in most children, so that they could practice reporting different kinds of strategies.

After this familiarisation stage, the child was asked to solve 15 problems presented on cards one at a time. The child was told to give the answer verbally as soon as (s)he knew it and to write the answer on the worksheet afterwards. No feedback was given on response accuracy. While the child was solving the problem, his/her behaviour was observed by the interviewer. Immediately after an answer was given, the interviewer asked, "*How did you solve this problem*?" No suggestion was made as to possible strategies so as not to influence the child's response. However, if the child gave a response that did not concur with the observed behaviour, the interviewer gave a single supplementary prompt based on that behaviour. The supplementary prompt was included because previous studies indicate that strategies that are more readily verbalisable may be reported at the expense of less verbalisable strategies such as retrieval, even when processing has been largely automatic (Ericsson & Simon, 1993; Kirk & Ashcraft, 2001). For example, one child solved the problem 5 × 5 in one second but reported using a laborious addition strategy: "*5 plus 5 is 10 and plus 5 is 15 and then 20 and then 25*." The interviewer prompted her by saying, "*Did you do all that just now*?" The child replied, "*No, I really just knew that 5 × 5 is 25*." In this case, the child reported using repeated addition when in fact she had retrieved the answer from memory.

For each problem, the interviewer recorded the child's responses in an individual interview booklet. Responses were recorded for both initial reports (i.e., responses to the initial question "*How did you solve this problem*?") and, if applicable, supplementary reports (i.e., responses to the supplementary prompt, if given) according to categories described in the section 'Variables'. The recordings of all interviews were subsequently analysed by the first author and children's responses were coded a second time. Inter-rater reliability was computed as Krippendorff's alpha coefficient for nominal data; the obtained coefficient of 0.86 indicates good reliability (Krippendorff, 2004).

2.3. Instruments

Fourteen single-digit multiplication problems and one problem involving a multiplicand of 10 were presented. Five problems involved multiples of 1, 2 or 5. These are treated in the Grade 2 curriculum and it was likely that these problems could be solved by direct retrieval by many Grade 3 children. Six problems involved multiples of 3, 4 or 6. These are included in the Grade 3 first semester curriculum and were familiar to the children. This allowed the possibility of a range of solution strategies – both retrieval and procedural – on the same problem. For example, 3 × 4 could be solved by retrieval, by using derived facts (e.g., 3 × 5 − 3), or by counting three series of four. Finally, three problems were included involving multiples of 7, 8 or 9. These problems are only treated in the second semester of Grade 3 and were unfamiliar to most children. These problems could therefore preferentially activate a procedural solution strategy rather than direct retrieval. The full list of problems used is given in the Appendix A; problems operationalised as ties and fives are also indicated.

2.4. Variables

Task-related variables were derived from the interview booklets and recordings. When no answer was given (i.e., 11 trials (0.7%)), the trial was treated as missing data and not analysed further. For each participant, the following variables were recorded.

Table 1
Number and mean age of participants across sex and math ability level.

	Math ability							
	Low		Below average		Average		Above average to high	
	N (%)	*y;m*	*N* (%)	*y;m*	*N* (%)	*y;m*	*N* (%)	*y;m*
Boys	8 (7.5)	9;0	10 (9.3)	8;9	20 (18.7)	8;11	9 (8.4)	8;8
Girls	13[a] (12.1)	9;1	19 (17.8)	9;1	17 (15.9)	8;8	11 (10.3)	8;9
Total	21 (19.6)	9;1	29 (27.1)	8;11	37 (34.6)	8;10	20 (18.7)	8;8

Note. [a]One participant was unresponsive from the start of her interview, which was then aborted. The participant was excluded from further analysis.

2.4.1. Reported strategy

Reported strategies were denoted as one of six types derived from previous research in this area (Cooney et al., 1988; LeFevre et al., 1996; Mabbott & Bisanz, 2003). These were: *repeated addition*, *derived facts*, *table string*, *retrieval*, *other*, or *unclassified*. Although *sequential counting* was initially included as a separate category, it was assimilated into the category *other* as it occurred on only 28 trials (1.8%), 12 of which were for a single participant. The classification *repeated addition* was used for statements as "4 plus 4 plus 4" or "4 plus 4 is 8 plus 4 is 12". This classification was also used for responses that began with repeated addition but finished with another addition strategy (e.g., solving 8×4 as $8 + 8 = 16$; $10 + 10 = 20$; $20 + 6 = 26$; $26 + 6 = 32$). Responses classified as *derived facts* involved using a known fact to derive a solution to the current problem (e.g., solving 3×4 as $2 \times 4 + 4$ or solving 5×8 as $10 \times 5 - 10$). *Table string* indicated when a memorised string was used by rote (e.g., reciting the string "5, 10, 15, 20, 25" to solve 5×5 or reciting "3 times 1 is 3, 3 times 2 is 6, 3 times 3 is 9, 3 times 4 is 12" to solve 3×4). Responses were classified as *retrieval* on the basis of statements as "I just knew it", "I remembered it" or "I know that by heart". The category *other* included various idiosyncratic or low-frequency strategies. The category *unclassified* was used when a child was unable to describe the strategy used and for responses that were too unclear to be assigned to another category. Initial and supplementary verbal reports were combined into a single reported strategy variable. Where a supplementary report had been solicited, this was used as the value of the variable; otherwise the initial report was used. Supplementary prompts were given to 45 children (42.1%) on 65 trials (4.1%). Nineteen children (17.8%) on 26 trials (1.6%) changed the strategy reported after prompting; of these, 16 (15%) changed their initial report to *retrieval* on 23 trials (1.4%).

2.4.2. Response latency

For each problem, response latency was measured as the number of seconds elapsed (rounded to the nearest hundredth of a second) between presentation of the problem and the child giving the answer verbally. Each response was measured by two coders independently using digital stopwatches and the average of these two measurements was used as the response latency value. Responses lasting longer than 30 seconds (i.e., 52 trials (3.3%)) were excluded from analysis, as memory traces were then considered to be less amenable to reliable reporting.

2.4.3. Response accuracy

For each problem presented, the answer given was coded as correct or incorrect. If the child gave an answer and corrected it immediately, the corrected answer was coded. If the answer was corrected later – when writing it down or explaining the strategy used – the initial answer was coded.

2.4.4. Error type

Error trials were classified according to typologies derived from previous research noted in the Introduction (Lemaire & Siegler, 1995; Siegler, 1988; Verguts & Fias, 2005). Errors were identified as *multiplicand-related*, *close miss* or *other* (all remaining errors). The *other* category included *tables-related* errors and *operation confusion* errors (i.e., the sum of the multiplicands rather than the product), which occurred on only 4 trials (0.3%).

2.4.5. Cognitive characteristics

Participating schools provided information on each child's math ability level, phonological decoding ability level, vocabulary level and reading comprehension level. Ability level was indicated by a grade in the range I (high) to V (low) on standardised, norm-referenced tests, with a grade of III representing the average level. These tests were provided by the Dutch Central Institute for Test Development (CITO). The math test comprised problems on numbers and number relations, simple arithmetic, proportions, measurement, geometry, money, and time. The phonological decoding ability test measured the speed and accuracy with which individual printed words were correctly decoded. The vocabulary test required students to choose the meaning, opposite meaning or synonym of indicated words. The reading comprehension test comprised passages of text accompanied by questions probing understanding of different elements of the text. Vocabulary and reading comprehension grades were averaged to give verbal ability level. All these variables were treated as continuous variables in the analyses, rescaled to start at a meaningful zero point. Correlations between math ability, speed of phonological decoding and verbal ability were all moderate to high: $r_{\text{math-pd}} = .29$, $N = 96$, $p = .005$; $r_{\text{math-verb}} = .54$, $N = 106$, $p < .001$; $r_{\text{pd-verb}} = .42$, $N = 96$, $p < .001$.

2.5. Analysis

Reported strategies *other* and *unclassified* and error type *other* were excluded from analysis, as there were no hypotheses for these. To account for correlated responses within children, multilevel models with random intercepts for children and problems were used. Problems were nested within children and problem characteristics and child cognitive characteristics were added as predictors at the corresponding levels.

To test patterns (P1) to (P4), analyses were performed using the lme4 package for the R statistical software environment (Bates, Maechler, & Bolker, 2010) with an α of .05. The standard Maximum

Table 2
Reported strategies: use, errors, accuracy and response latencies.

	Trials						Participants	
				Latency (seconds)				
Reported strategy	Use[a] (%)	Errors[b]	Accuracy[c]	Mean	Median	*SD*	Use[d] (%)	Errors[e] (%)
Repeated addition	178 (*11.7*)	23 (*12.9*)	87.1%	6.03	3.05	6.39	72 (*67.9*)	9 (*8.5*)
Derived facts	531 (*34.8*)	49 (*9.2*)	90.8%	6.92	4.89	5.97	97 (*91.5*)	32 (*30.2*)
Table string	73 (*4.8*)	7 (*9.6*)	90.4%	8.86	6.23	6.97	35 (*33.0*)	7 (*6.6*)
Retrieval	622 (*40.7*)	10 (*1.6*)	98.4%	2.44	1.91	1.97	101[f] (*95.3*)	8 (*7.5*)
Other	59 (*3.9*)	13 (*22.0*)	78.0%	8.48	3.50	8.24	30 (*28.3*)	7 (*6.6*)
Unclassified	64 (*4.2*)	11 (*17.2*)	82.8%	5.98	2.74	6.67	30 (*28.3*)	8 (*7.5*)
Total	1527 (*100*)	113 (*7.4*)	92.6%	5.10	2.79	5.54	106 (*100*)	52 (*49.1*)

Notes. [a]Number and percentage of valid trials (RT ≤ 30) for which the strategy was reported.
[b] Number and percentage of errors on valid trials for which the strategy was reported.
[c] Percentage correct of valid trials for which the strategy was reported.
[d] Number and percentage of participants who reported using the strategy at least once.
[e] Number and percentage of participants who made errors using the reported strategy.
[f] Seven participants used retrieval only on the problem involving the multiplicand 1.

Table 3
Overview of the estimated models with response latency as dependent variable.

Model	Nested model	Fixed effects[a]	Deviance	AIC	BIC	- Log Likelihood	Model *df*	LR test[b]	
								df	Λ
M0		Intercept	8892	8900	8921	4446	4		
M1	M0	+ strategy	7950	7964	8001	3975	7	3	941.92***
M2	M1	+ product	7945	7961	8003	3972	8	1	5.08*
M3	M2	+ ties	7945	7963	8010	3972	9	1	0.22
M4	M2	+ fives	7936	7954	8001	3968	9	1	8.56**
M5	M4	+ math ability	7935	7955	8007	3967	10	1	1.69
M6	M4	+ strategy × math	7929	7955	8023	3964	13	4	7.33
M7	M4	+ ph.dec.speed	7154	7174	7226	3577	10	1	781.90***
M8	M7	+ strategy × pds	7133	7159	7226	3567	13	3	21.23***
M9	M8	+ verbal ability	7132	7160	7232	3566	14	1	0.71
M10[c]	**M8**	**+ strategy × verb**	**7109**	**7143**	**7230**	**3554**	**17**	**4**	**24.30*****

Notes. [a] Random effects over persons and problems are captured in their respective intercepts.
[b] The LR (likelihood ratio) test comprises a comparison between the model and the nested model.
[c] Best fitting model with intercept set to 0: persons (VAR = 4.12, SD = 2.03); problems (VAR = 1.80, SD = 1.34).
Best fitting model is in bold font.
***p<.001, **p<.01, *p<.05.

Likelihood method of estimating free parameters was used when the outcome variable was continuous (i.e., response latency). The Laplace Approximation estimation method was used when the outcome variable was dichotomous (i.e., accuracy and error type). Model fit was evaluated with the Akaike Information Criterion (AIC), models with lower AIC values being preferred.

The relation between reported strategy and response latency was tested with a multilevel linear mixed model with reported strategy as predictor and response latency as dependent variable. Problem characteristics were then included as predictors: problem size was operationalised as the product of the problem multiplicands and the hypothesised mitigating effect of ties and fives was indicated by two dichotomous variables (0 = non-tie/non-five; 1 = tie/five). Response latencies between reported strategies were compared using simple contrasts. The model was then extended to include child cognitive characteristics (i.e., math ability, speed of phonological decoding, verbal ability). These were sequentially added as predictors, as were their interactions with reported strategy. Only significant effects were retained in each subsequent model. Explained variance of the final model was calculated using a method described by Xu (2003) for measuring explained variation in linear mixed effects models. With this method, a partial coefficient measure (r^2) is calculated as:

$$1-\left(estimated\ \sigma^2/estimated\ {\sigma_0}^2\right)$$

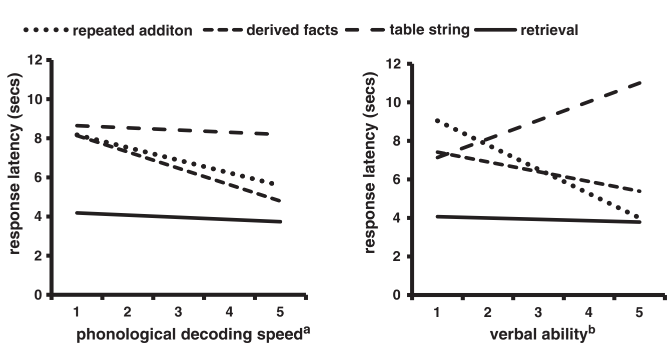

Fig. 1. Interactions between reported strategy and (a) phonological decoding speed (b) verbal ability. Note. Graphs are estimated at average problem size (i.e., 21). [a] Latency values are calculated as: strategy estimate + (problem size estimate*average problem size) + (phonological decoding speed estimate*phonological decoding speed value) + (strategy × phonological decoding speed estimate*phonological decoding speed value) (see Model M8; fives are not applicable at problem size = 21). [b] Latency values are calculated as: strategy estimate + (problem size estimate*average problem size) + (phonological decoding speed estimate*verbal ability value) + (strategy × phonological decoding speed estimate*verbal ability value) + (strategy × verbal ability estimate*verbal ability value) (see Model M10; fives are not applicable at problem size = 21).

where *estimated* σ^2 is the amount of variation in the dependent variable that is not explained by the fixed and random effects and *estimated* σ_0^2 is the amount of variation in the null model.

To test the relation between reported strategy and accuracy, the analyses were repeated using a multilevel binary logistic model with accuracy as the dependent variable (0 = incorrect; 1 = correct). To test the relation between reported strategy and error type, a multilevel binary logistic model was used to determine which factors influence the types of error produced (0 = multiplicand-related; 1 = close miss). Explanatory power of these models was calculated as Tjur's (2009) coefficient of discrimination. This is the difference in the averages of fitted values for 'successes' and 'failures' (in terms of a binary outcome) and represents the model's ability to discriminate between these outcomes. Coefficient values close to 0 indicate that the model has little explanatory power and values close to 1 indicate near perfect fit.

Pattern (P5) could not be tested in the above way as not every strategy was sufficiently frequently reported for each problem for the model to converge. Therefore, a mixed models ANOVA including random effects for individuals was performed in IBM SPSS Statistics 20®, with planned contrasts to compare problem size between strategies after controlling for ties and fives.

3. Results

3.1. Strategy use

For each reported strategy, Table 2 presents (a) frequency of use, errors, accuracy and response latencies across all valid trials (i.e., excluding trials with RT >30 and non-response trials); (b) the number and percentage of participants who reported using the strategy at least once and the number and percentage of participants who made errors when using the strategy. Considerable diversity in strategy use within individuals was reported. Three participants (2.8%) reported using only one strategy; 11 (10.4%) reported two different strategies, 43 (40.6%) reported three, 35 (33.0%) reported four, 13 (12.3%) reported five and one participant (0.9%) reported six different strategies.

3.2. Reported strategies and response latencies

An overview of the estimated models is presented in Table 3. Strategy was a significant predictor of response latency (M1; $\chi^2(3) = 941.92$, $p < .001$) when compared to a null model (M0) comprising only random intercepts. Prediction improved by including problem size (M2; $\chi^2(1) = 5.08$, $p = .02$) and fives (M4; $\chi^2(1) = 8.56$, $p = .003$) as covariates but including ties (M3) did not improve the model. An increase of 1 in problem size was estimated to lead to 0.12 seconds increase in

Table 4
Overview of the estimated models with accuracy as dependent variable.

Model	Nested model	Fixed effects[a]	Deviance	AIC	BIC	- Log Likelihood	Model *df*	LR test[b] *df*	LR test[b] Λ
M0		Intercept	637	643	659	318	3		
M1	M0	+ strategy	504	516	548	252	6	3	132.44***
M2	M1	+ product	496	510	546	248	7	1	8.64**
M3	M2	+ ties	495	511	553	248	8	1	0.50
M4	M2	+ fives	492	508	550	246	8	1	3.52†
M5	M4	+ math ability	486	504	551	243	9	1	6.53*
M6	M5	+ strategy × math	483	507	570	242	12	3	2.24
M7	**M5**	**+ ph.dec.speed**	**441**	**461**	**512**	**220**	**10**	**1**	**45.05*** **
M8	M7	+ strategy × pds	439	465	532	220	13	3	1.35
M9	M7	+ verbal ability	441	463	519	220	11	1	0.02
M10	M7	+ strategy × verb	437	465	537	218	14	4	3.65

Notes. [a] Random effects over persons and problems are captured in their respective intercepts.
[b] The LR (likelihood ratio) test comprises a comparison between the model and the nested model.
Best fitting model is in bold font.
***$p < .001$, **$p < .01$, *$p < .05$, †$p = .06$.

latency ($B = 0.12$, $SE = 0.03$), but if the problem was a five, latency decreased by nearly three seconds ($B = -2.83$, $SE = 0.83$). These effects (i.e., larger problems take longer to solve unless they involve a tie or a five) were largely as expected; only the mitigating effect of ties was not found. It should be noted, though, that only two ties were included: 5 × 5 (which is also a five) and 6 × 6. Thus, it is likely that there were not enough tie measurements to detect this effect. The fixed effects of the strategies on response latency in increasing order were: *retrieval* ($B = 1.43$, $SE = 0.83$), *derived facts* ($B = 4.21$, $SE = 0.85$), *repeated addition* ($B = 4.63$, $SE = 0.87$) and *table strings* ($B = 6.11$, $SE = 0.96$). Simple contrasts revealed significant differences between each pair of strategies except *repeated addition* and *derived facts* ($p = .29$). Thus, as expected, *retrieval* was faster than the procedural strategies but, though *repeated addition* was slower than *derived facts*, this difference was not significant.

Extending the model to include cognitive characteristics showed that neither math ability (M5) nor its interaction with reported strategy (M6) improved the model. This means that individual differences in math ability were not relevant to the relation between reported strategies and response latency. Both speed of phonological decoding (M7; $\chi^2(1) = 781.90$, $p < .001$) and its interaction with strategy (M8; $\chi^2(3) = 21.23$, $p < .001$) did improve the model. Children with higher phonological decoding speed generally had lower response latencies ($B = -0.65$, $SE = 0.25$). The interaction with strategy showed that this effect was strongest in children who reported using procedural (i.e., *repeated addition* or *derived facts*) strategies. Verbal ability did not influence response latency in general (M9), but the interaction between verbal ability and reported strategy was a significant predictor over and above the effects of phonological decoding speed (M10; $\chi^2(4) = 24.30$, $p < .001$). When children reported using *repeated addition*, those with higher verbal ability had faster solution times than children with lower verbal ability, as expected. However, when children reported using a *table string*, those with lower verbal ability were faster than those with higher verbal ability. A posthoc ANOVA confirmed that this was not due to children with lower verbal ability using table strings on smaller and easier problems than children with higher verbal ability ($F(7,65) = 0.73$, $p = .65$).

Explained variance of the final model was 19.35%, a medium-sized effect (Cohen, 1992). Fig. 1 shows the interaction effects. Note that adding effects in a different order within each set of predictors (e.g., adding fives before ties) did not affect the substantive conclusions.

3.3. Reported strategies and accuracy

An overview of the estimated models is presented in Table 4. On the whole, children's solutions were highly accurate (92.6%; see Table 2). Strategy was a significant predictor of accuracy (M1; $\chi^2(3) = 132.44$, $p < .001$) when compared to a null model (M0) comprising only random intercepts. Accuracy prediction improved by including problem size as a covariate (M2; $\chi^2(1) = 8.64$, $p = .003$) and there was a strong tendency for fives (M4; $\chi^2(1) = 3.52$, $p = .06$) but not ties (M3) to improve the model. The odds[4] that an average child solved a problem correctly decreased by .10 when problem size increased by 1 ($B = -0.10$, $SE = 0.03$); if the problem was a five then the odds of solving it correctly increased by 4.39 ($B = 1.48$, $SE = 0.76$). These effects are in the expected direction. The fixed effects of the strategies on accuracy in increasing order were: *repeated addition* ($B = 4.89$, $SE = 0.87$), *table strings* ($B = 5.61$, $SE = 1.00$), *derived facts* ($B = 6.27$, $SE = 0.91$) and *retrieval* ($B = 7.51$, $SE = 0.95$). Simple contrasts indicated that *derived facts* and *retrieval* were more accurate than *repeated addition*, as expected. Although *retrieval* was more accurate than *derived facts* and *table strings*, these differences were not significant.

Regarding cognitive characteristics, math ability improved the model (M5; $\chi^2(1) = 6.53$, $p = .01$), with higher ability children being more likely to solve problems correctly ($B = 0.49$, $SE = 0.19$, $OR = 1.63$), as expected. Speed of phonological decoding also improved model fit (M7; $\chi^2(1) = 45.05$, $p < .001$); however, its effect on correct solution was not significant ($B = -0.07$, $SE = 0.12$, $p = .58$). Verbal ability (M9) did not influence accuracy and none of the cognitive characteristics interacted with reported strategy (M6, M8, M10). Predictor order did not affect these conclusions. Tjur's coefficient of discrimination was 0.28, indicating moderate model fit.

Table 5
Frequency and percentage of error types per reported strategy.

Reported strategy	Error type[a]: Multiplicand-related (%)	Close miss (%)	Other[b] (%)	Total (%)
Repeated addition	9 (*8.0*)	6 (*5.3*)	8 (*7.1*)	23 (*20.4*)
Derived facts	21 (*18.6*)	17 (*15.0*)	11 (*9.7*)	49 (*43.4*)
Table string	6 (*5.3*)	1 (*0.9*)	0	7 (*6.2*)
Retrieval	9 (*8.0*)	1 (*0.9*)	0	10 (*8.8*)
Other	9 (*8.0*)	1 (*0.9*)	3 (*2.7*)	13 (*11.5*)
Unclassified	4 (*3.5*)	2 (*1.8*)	5 (*4.4*)	11 (*9.7*)
Total	58 (*51.3*)	28 (*24.8*)	27 (*23.9*)	113 (*100*)

Notes. [a] Number and percentage of error trials ($N = 113$) for which the error type was reported.
[b] Including 2 *tables-related* and 2 *operation confusion* errors.

[4] The odds of a positive outcome for a particular predictor is $\exp(B_{pred})$. When the odds are <1 the likelihood of obtaining a positive outcome decreases for each unit increase in the predictor; when the odds are >1, a positive outcome is more likely.

Table 6
Overview of the estimated models with error type as dependent variable.

Model	Nested model	Fixed effects[a]	Deviance	AIC	BIC	- Log Likelihood	Model *df*	LR test[b]	
								df	Λ
M0		Intercept	108	114	121	54	3		
M1	M0	+ strategy	85	97	110	42	6	3	22.93***
M2	M1	+ product	85	99	114	42	7	1	0.12
M3	M1	+ ties	85	99	114	42	7	1	0.06
M4	M1	+ fives	84	98	113	42	7	1	0.98
M5	M1	+ math ability	84	98	114	42	7	1	0.38
M6	**M1**	**+ ph.dec.speed**	**68**	**82**	**97**	**34**	**7**	**1**	**16.54*** **
M7	M6	+ verbal ability	68	84	101	34	8	1	0.08

Notes. [a] Random effects over persons and problems are captured in their respective intercepts.
[b] The LR (likelihood ratio) test comprises a comparison between the model and the nested model.
Interactions with strategy are not shown because of lack of convergence due to sparsity of errors.
Best fitting model is in bold font.
***$p < .001$.

3.4. Reported strategies and error type

Table 5 shows the frequency and percentage of error types per reported strategy and an overview of the estimated models is presented in Table 6. Strategy was a significant predictor of error type (M1; $\chi^2(3) = 22.93, p < .001$) when compared to a null model (M0) comprising only random intercepts. Prediction was not improved by including problem characteristics (M2, M3, M4). As expected, children were more likely to commit multiplicand-related errors than 'close misses' on *retrieval* ($B = -2.23$, $SE = 1.07$, $OR = 0.11$). Although *repeated addition* ($B = -0.40$, $SE = 0.54$, $OR = 0.67$), *derived facts* ($B = -0.22$, $SE = 0.33$, $OR = 0.80$) and *table strings* ($B = -1.81$, $SE = 1.09$, $OR = 0.16$) were also more likely to have multiplicand-related errors, these effects were not significant. This is likely due to the small number of errors per strategy and error type (see Table 5). Regarding cognitive characteristics, only phonological decoding speed (M6; $\chi^2(1) = 16.54$, $p < .001$) improved model fit; however, its effect on correct solution was not significant ($B = 0.27$, $SE = 0.21$, $p = .20$). Predictor order did not affect these conclusions. Tjur's coefficient of discrimination was 0.18, a medium-sized effect.

3.5. Strategy choice and problem characteristics

Problem size ranged from 3 to 54. Taking account of correlated responses and problem characteristics, problem size differed between strategies ($F(3,455) = 71.20$, $p < .001$, $\eta_p^2 = .32$), a large effect (see Fig. 2). Planned contrasts indicated that problem size was smaller when *repeated addition* ($M = 16.80$, $SD = 9.87$) and *retrieval* were reported ($M = 21.94$, $SD = 13.65$) and larger when *table strings* ($M = 28.03$, $SD = 12.48$) and *derived facts* ($M = 29.11$, $SD = 12.51$) were reported. *Retrieval* was the most frequent strategy used on fives and *table string* use peaked at larger fives. These results conform to expectation except that *repeated addition* was expected to be used more often on larger problems; however it was generally more often employed on smaller problems.

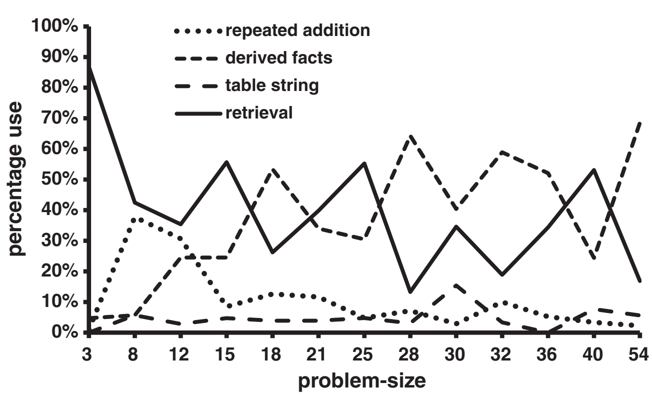

Fig. 2. Reported strategy use as a function of problem size.

4. Discussion

The purpose of this study was to establish whether verbal reports accurately represent third graders' thinking processes (i.e., veridicality) for simple multiplication problem-solving. Of course, as others have argued (e.g., Russo et al., 1989), there is no absolute criterion against which to judge the veridicality of children's verbal reports. In this study, the approach taken was to assess the degree to which verbal reports predict children's problem-solving performance in ways that correspond to known patterns of response latency, accuracy, errors and strategy choice, taking account of relevant problem characteristics and cognitive characteristics that are known to affect the relation between strategy use and multiplication performance. With this approach, a close correspondence with known patterns would demonstrate veridicality. An important asset of this study compared to earlier research, is that correlated responses within individuals are taken account of through multilevel techniques, thereby avoiding the problem of misleading results when data is aggregated across subjects or strategies (Cooney et al., 1988; Siegler, 1987, 1989).

4.1. Are verbal reports of children's multiplication strategy use veridical?

Verbal reports were found to predict children's problem-solving performance in ways that are largely consistent with the literature. Verbal reports corresponded to almost all of the expected patterns regarding strategies and performance (P1), problem characteristics (P2), cognitive characteristics (P3), error types (P4) and strategy choice versus problem characteristics (P5). Though certain effects were not significant in the sample, this is likely partly due to the fact that group sizes were relatively small when different levels of certain predictors (e.g., cognitive characteristics) were compared.

It is notable that procedural strategies were not fully distinguishable through the verbal reports: although use of derived facts was faster and more accurate than repeated addition as expected, the difference in solution speed was not significant. Moreover, two unexpected results were obtained: when table strings were reported, children with lower verbal ability were faster than children with higher verbal ability; and repeated addition was reported on smaller, rather than larger problems.

What do these discrepancies signify about the veridicality of third graders' verbal reports of multiplication strategy use? In answering this question, it is important to consider whether these results are compatible with what is known about children's multiplication learning. For one, third graders may not yet be highly fluent in the use of derived facts and may use shortcuts that are as procedurally demanding and time-consuming as addition strategies (Cooney & Ladd, 1992; Sherin & Fuson, 2005). Thus, solution speed when using these strategies may not differ much at this stage of multiplication learning. Also, the derived facts strategy was reported mainly on larger problems, while repeated addition was reported mainly on smaller problems, so that the speed advantage of derived facts over repeated addition could have been attenuated by the problem-size effect described earlier.

Second, the finding that children with lower verbal ability were faster than children with higher verbal ability when using table strings is compatible with the notion that use of table strings depends on recall from verbal memory, as noted earlier (Dehaene, 2011; Geary, 2004). Children with higher verbal ability may use a more elaborated – and therefore more time – consuming – table string recall (e.g., “one times five is five, two times five is ten, three times five is fifteen”), while children with lower verbal ability may use a more abbreviated form (e.g., “five, ten, fifteen”).

Third, the finding that repeated addition was reported on smaller, rather than larger problems could be an artefact of the instructional approach in the country of study (i.e., the Netherlands). Early multiplication learning with smaller problems emphasises repeated addition, while later instruction emphasises the use of ‘shortcut’ strategies (i.e., derived facts) such as the commutative rule, doubling, halving, ‘one-more’ and ‘one-less’, which are particularly handy for solving larger problems (Van Zanten, Barth, Faarts, Van Gool, & Keijzer, 2009). Children in the present study, who were at an intermediate stage of multiplication learning, may thus have adhered to the strategy/problem size combinations they encountered in their lessons, that is, repeated addition on smaller problems and derived facts on larger problems.

In short, though verbal reports did not fully predict children's multiplication performance in the expected ways, the results are largely consistent with what is known about children's multiplication learning and performance. This indicates that third graders' verbal reports of multiplication strategy use are sufficiently veridical to be used as a screening instrument in regular classroom settings when problem characteristics and cognitive characteristics are taken into account. With regard to identifying children who lag behind their peers, it also appears that mature strategies (i.e., retrieval) are not over-reported; rather, almost all discrepancies between initial reports and observed behaviour involved under-reporting of retrieval. Nonetheless, we recognise that a substantial amount of variance in performance remains unexplained, so that verbal reports as solicited here cannot be considered to be a completely secure method of identifying strategy use. Thus, improvements to the present procedure for soliciting verbal reports are indicated, as will be discussed later.

4.2. The importance of verbal reports for research into multiplication learning

This study has important implications for research into children's multiplication learning. Specifically, as others have argued for different reasons, assumptions about strategy use should not be based on response latencies alone. In this study, some children reported using laborious procedural strategies while their responses were extremely fast. For instance, one child reported solving the problem 7 × 3 as “*10 times 3 is 30 so then 9 × 3 is 3 less, that's 27, and 8 × 3 is another 3 less and 7 × 3 another 3 less*”. Although she had solved the problem in only one second, she maintained: “*I can do all that really quickly*”. One might think that this child described a procedural strategy because she could not verbalise a faster memory-based strategy (Ericsson & Simon, 1993; Kirk & Ashcraft, 2001). However, the same child stated “*I've already memorised this*” for four other problems, showing that she was able to both recognise and report memory-based strategies.

Conversely, several children reported using retrieval but had long response latencies. Although it is possible that they were actually using a procedural strategy, another explanation should be considered. It is generally assumed that arithmetic facts are stored as networks of problem-answer associations in long-term memory and that activation spreads to adjacent nodes upon presentation of a particular problem (Galfano, Rusconi, & Umiltà, 2003; Verguts & Fias, 2005). For example, presentation of the problem 4 × 7 activates the problems 3 × 7, 5 × 7, 4 × 6 and 4 × 8 and the close operand-related answers 21, 35, 24 and 32. As facts become learned, the associative strength of the correct answer comes to exceed that of competing answers, increasing the likelihood of that answer being retrieved from memory (Siegler & Shipley, 1995). However, when facts are not yet fully learned and several candidates are activated, children need to decide which answer to give. Two examples are illustrative. One participant stated: “*I've got lots of answers in my head and then I think: yes, that one belongs to this sum, and then I know it*”. Another stated: “*Then I saw all the tables before me again and then I saw that table and there was 5 × 6 and then I thought about it again and then my brain said that it was actually just 30*”. Such decision-making processes – possibly even involving visualisation of multiplication arrays – could be quite time-consuming, yet is likely that children experience this as a form of retrieval. Thus, although it is often assumed that longer latencies indicate the use of procedural strategies, it may well be that longer latencies are also produced when answers are retrieved but accompanied by further decision-making processes (cf. the ‘confidence criterion’ described by Siegler and collegues; Siegler, 1988; Siegler & Shipley, 1995). Together, these findings support the necessity of including verbal reports in the assessment of children's cognitive processing.

4.3. Educational implications

As mainstream primary schools become increasingly inclusive, it is important to equip teachers with the means to detect children who lag behind their peers in thinking processes that impact their school performance (Humphrey et al., 2006; UNESCO, 1994). This study provides an interview-based instrument that could be used by teachers in regular classroom settings to help determine which strategies children use to solve simple multiplication problems. The influence of child cognitive characteristics was investigated using scores from standardised, norm-referenced tests of a type that is widely used in primary schools and with which teachers are familiar. This increases the accessibility of the methods used for classroom practice.

The results show that third graders' verbal reports in response to the interview protocol give a reasonable indication of their actual mental processing during simple multiplication problem-solving. Nonetheless, at an intermediate stage of multiplication learning with instructional methods that emphasise a specific progression in strategy development, different procedural strategies are not completely distinguishable through the type of verbal report solicited here. Furthermore, discrepancies between children's initial responses and their observed behaviour led to 18% of the children changing at least one verbal report, and almost a third were unable to report clearly what strategy they used on at least one problem. In practice, this means that teachers need to observe children's behaviour during problem-solving and probe in more detail when behaviour and reports appear discrepant, when children report using a procedural strategy, and when children are unable to indicate what strategy they have used. As discussed later, future research is necessary to establish how teachers can reliably obtain this information.

When teachers can assess children's progression in strategy use, they can adapt instruction to individual levels of expertise. This principle stems from the finding that instructional methods that are effective for novice learners may be counterproductive for more expert learners and vice versa (the expertise reversal effect; Kalyuga, 2007). To illustrate: children in this study – who, as noted, are at an intermediate stage of learning multiplication – reported considerable diversity in strategy use, consistent with previous research. This is in context of the instructional approach previously described, where use of derived facts and then retrieval should eventually come to replace addition-based strategies. It is notable, therefore, that 68% of children in this study still reported using repeated addition, primarily on smaller problems. This suggests that some children may persist in using the strategies that they first learned on the problems for which they learned them. Persistence in using more primitive strategies can be a sign of mathematics difficulties (Geary, 2004; Geary et al., 2012), while persistence in using a strategy that has worked well in the past may be a sign of lower cognitive flexibility, which is also associated with lower math performance (Bull & Scerif, 2001). These kinds of insights can alert teachers to particular problems that individual students may be having and help them to adapt their approach accordingly.

4.4. Future research

The present study used a strictly proscribed interview protocol to ensure that children's responses were not influenced by variations in instructions (Kirk & Ashcraft, 2001; Taylor & Dionne, 2000) and to safeguard the generalisability of the results. Consequently, the protocol could not fully detect individual variations in cognitive processing when procedural strategies were reported. As this information is necessary for determining where an individual is in his/her learning trajectory as well as difficulties that (s)he may experience in appropriating more advanced strategies, future research that indicates how this information can be reliably and validly obtained is necessary. Thus, interview protocols should be developed that could be used by teachers to probe children's thinking more deeply, perhaps using simplified versions of protocols used by professional diagnosticians as a point of departure. This should be accompanied by initiatives to increase teachers' professional knowledge about children's problems and misconceptions in math learning, so that teachers are better equipped to respond to difficulties that they uncover in this way.

Certain aspects of the study design could also be improved in future research. First, we used only one measure of phonological processing, namely a routinely administered phonological decoding test. Given the importance of other aspects of phonological processing on multiplication performance, as discussed, it would be of interest to investigate the relation between these aspects and verbal reports directly using a wider range of tests. Furthermore, other factors not examined here could also affect the veridicality of verbal reports and account for some of the unexplained variance in the investigated models. These could include children's verbal fluency, their conformance to social desirability or class norms, the quality of the teacher-student relationship, as well as characteristics of the instructional methods used in the classroom. These aspects could also be included in future research. Finally, the veridicality of verbal reports at a later stage of multiplication learning (e.g., Grade 4) and in systems using other instructional methods should be investigated, so that developmental trends in relation to didactical approach can be explored.

Acknowledgements

The authors thank Michelle Gemmink and the children, parents and teachers of the participating schools for their contributions to this study. This research was partly supported by a grant from the Dutch National Initiative Brain and Cognition (NIBC; project number 056-31-013).

Appendix A

The following problems were used in the interviews:

- One problem involving a multiple of 10: 4×10
- Five problems involving multiples of 1, 2 or 5: $3 \times 1, 4 \times 2, 6 \times 2, 3 \times 5, 5 \times 5$
- Six problems involving multiples of 3, 4 or 6: $6 \times 3, 7 \times 3, 3 \times 4, 8 \times 4, 5 \times 6, 6 \times 6$
- Three problems involving multiples of 7, 8 or 9: $4 \times 7, 5 \times 8, 6 \times 9$

Ties: 5×5; 6×6.
Fives: $4 \times 10, 3 \times 5, 5 \times 5, 5 \times 6, 5 \times 8$. Though the problem 4×10 does not include 5 as a multiplicand, its solution is a multiple of 5 and performance on this problem was similar to that on the other fives.

References

Alibali, M. W. (1999). How children change their minds: Strategy change can be gradual or abrupt. *Developmental Psychology*, *35*, 127–145.

Bates, D., Maechler, M., & Bolker, B. (2010). Lme4: Linear mixed-effects models using S4 classes [Computer software]. Retrieved from http://CRAN.R-project.org/package=lme4.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning. *Journal of Experimental Psychology: General*, *136*, 539–550.

Bull, R., & Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: Inhibition, switching, and working memory. *Developmental Neuropsychology*, *19*, 273–293.

Chin, J. M., & Schooler, J. W. (2008). Why do words hurt? Content, process, and criterion shift accounts of verbal overshadowing. *European Journal of Cognitive Psychology*, *20*, 396–413.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 115–159.

Cooney, J. B., & Ladd, S. F. (1992). The influence of verbal protocol methods on children's mental computation. *Learning and Individual Differences*, *4*, 237–257.

Cooney, J. B., Swanson, H. L., & Ladd, S. F. (1988). Acquisition of mental multiplication skill: Evidence for the transition between counting and retrieval strategies. *Cognition and Instruction*, *5*, 323–345.

Crutcher, R. J. (1994). Telling what we know: The use of verbal report methodologies in psychological research. *Psychological Science*, *5*, 241–244.

De Brauwer, J., Verguts, T., & Fias, W. (2006). The representation of multiplication facts: Developmental changes in the problem size, five, and tie effects. *Journal of Experimental Child Psychology*, *94*, 43–56.

De Smedt, B., Taylor, J., Archibald, L., & Ansari, D. (2010). How is phonological processing related to individual differences in children's arithmetic skills? *Developmental Science*, *13*, 508–520.

Dehaene, S. (2011). *The number sense: How the mind creates mathematics* (2nd ed.). New York, NY: Oxford University Press.

Dowker, A. (2004). What works for children with mathematical difficulties? *Research Report No. 554*. Nottingham: U.K. Department for Education and Skills.

Dubé, A. K., & Robinson, K. M. (2010). Accounting for individual variability in inversion shortcut use. *Learning and Individual Differences*, *20*, 687–693.

Durand, M., Hulme, C., Larkin, R., & Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *Journal of Experimental Child Psychology*, *91*, 113–136.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Revised ed.). Cambridge, MA: MIT Press.

Fox, M. C., Ericsson, K. A., & Best, R. (2011). Do procedures for verbal reporting of thinking have to be reactive? A meta-analysis and recommendations for best reporting methods. *Psychological Bulletin*, *137*, 316–344.

Foxman, D., & Beishuizen, M. (2002). Mental calculation methods used by 11-year-olds in different attainment bands: A reanalysis of data from the 1987 APU survey in the UK. *Educational Studies in Mathematics*, *51*, 41–69.

Fuchs, L. S., Fuchs, D., Compton, D. L., Powell, S. R., Seethaler, P. M., Capizzi, A. M., et al. (2006). The cognitive correlates of third-grade skill in arithmetic, algorithmic computation, and arithmetic word problems. *Journal of Educational Psychology*, *98*, 29–43.

Galfano, G., Rusconi, E., & Umiltà, C. (2003). Automatic activation of multiplication facts: Evidence from the nodes adjacent to the products. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *56*, 31–61.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, *37*, 4–15.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2012). Mathematical cognition deficits in children with learning disabilities and persistent low achievement: A five-year prospective study. *Journal of Educational Psychology*, *104*, 206–223.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties. *Journal of Learning Disabilities*, *38*, 293–304.

Grabner, R. H., Ansari, D., Koschutnig, K., Reishofer, G., Ebner, F., & Neuper, C. (2009). To retrieve or to calculate? Left angular gyrus mediates the retrieval of arithmetic facts during problem solving. *Neuropsychologia*, *47*, 604–608.

Grabner, R. H., & De Smedt, B. (2011). Neurophysiological evidence for the validity of verbal strategy reports in mental arithmetic. *Biological Psychology*, *87*, 128–136.

Hecht, S. A., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2001). The relations between phonological processing abilities and emerging individual differences in mathematical computation skills: A longitudinal study from second to fifth grades. *Journal of Experimental Child Psychology*, *79*, 192–227.

Humphrey, N., Bartolo, P., Ale, P., Calleja, C., Hofsaess, T., Janikova, V., et al. (2006). Understanding and responding to diversity in the primary classroom: an international study. *European Journal of Teacher Education*, *29*, 305–318.

Imbo, I., & Vandierendonck, A. (2007). Do multiplication and division strategies rely on executive and phonological working memory resources? *Memory & Cognition*, *35*, 1759–1771.

Imbo, I., & Vandierendonck, A. (2008a). Effects of problem size, operation, and working-memory span on simple-arithmetic strategies: differences between children and adults? *Psychological Research*, *72*, 331–346.

Imbo, I., & Vandierendonck, A. (2008b). Practice effects on strategy selection and strategy efficiency in simple mental arithmetic. *Psychological Research*, *72*, 528–541.

Jordan, N. C., Hanich, L. B., & Kaplan, D. (2003). Arithmetic fact mastery in young children: A longitudinal investigation. *Journal of Experimental Child Psychology*, *85*, 103–119.

Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review*, *19*, 509–539.

Kilpatrick, J., Swafford, J., & Findell, B. (2001). *Adding it up: Helping children learn mathematics.* Washington, DC: National Academy Press.

Kirk, E. P., & Ashcraft, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 157–175.

Krippendorff, K. (2004). Reliability in content analysis: Some common misconceptions and recommendations. *Human Communication Research*, *30*, 411–433.

LeFevre, J. -A., Bisanz, J., Daley, K. E., Buffone, L., Greenham, S. L., & Sadesky, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General*, *125*, 284–306.

LeFevre, J. -A., Fast, L., Skwarchuk, S. -L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, *81*, 1753–1767.

Lemaire, P., & Siegler, R. S. (1995). Four aspects of strategic change: Contributions to children's learning of multiplication. *Journal of Experimental Psychology: General*, *124*, 83–97.

Lidstone, J. S. M., Fernyhough, C., Meins, E., & Whitehouse, A. J. O. (2009). Inner speech impairment in children with autism is associated with greater nonverbal than verbal skills. *Journal of Autism and Developmental Disorders*, *39*, 1222–1225.

Lidstone, J. S. M., Meins, E., & Fernyhough, C. (2012). Verbal mediation of cognition in children with specific language impairment. *Development and Psychopathology*, *24*, 651–660.

Mabbott, D. J., & Bisanz, J. (2003). Developmental change and individual differences in children's multiplication. *Child Development*, *74*, 1091–1107.

Neil, G. J., & Higham, P. A. (2012). Implicit learning of conjunctive rule sets: An alternative to artificial grammars. *Consciousness and Cognition*, *21*, 1393–1400.

Ostad, S. A. (2013). Private speech use in arithmetical calculation: Contributory role of phonological awareness in children with and without mathematical difficulties. *Journal of Learning Disabilities*, *46*, 291–303.

Ostad, S. A., & Sorensen, P. M. (2007). Private speech and strategy use patterns: Bidirectional comparisons of children with and without mathematical difficulties in a developmental perspective. *Journal of Learning Disabilities*, *40*, 2–14.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, *20*, 110–122.

Robinson, K. M. (2001). The validity of verbal reports in children's subtraction. *Journal of Educational Psychology*, *93*, 211–222.

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition*, *17*, 759–769.

Sanchez, D. J., Gobel, E. W., & Reber, P. J. (2010). Performing the unexplainable: Implicit task performance reveals individually reliable sequence learning without explicit knowledge. *Psychonomic Bulletin & Review*, *17*, 760–796.

Sherin, B., & Fuson, K. (2005). Multiplication strategies and the appropriation of computational resources. *Journal for Research in Mathematics Education*, *36*, 347–395.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250–264.

Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, *117*, 258–275.

Siegler, R. S. (1989). Hazards of mental chronometry: An example from children's subtraction. *Journal of Educational Psychology*, *81*, 497–506.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In T. J. Simon, & G. S. Halford (Eds.), *Developing cognitive competence: New approaches to process modeling* (pp. 31–76). Hillsdale, NJ: Lawrence Erlbaum.

Siegler, R. S., & Stern, E. (1998). Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology: General*, *127*, 377–397.

Simmons, F. R., & Singleton, C. (2008). Do weak phonological representations impact on arithmetic development? A review of research into arithmetic and dyslexia. *Dyslexia*, *14*, 77–94.

Steel, S., & Funnell, E. (2001). Learning multiplication facts: A study of children taught by discovery methods in England. *Journal of Experimental Child Psychology*, *79*, 37–55.

Taylor, K. L., & Dionne, J. -P. (2000). Assessing problem-solving strategy knowledge: The complementary use of concurrent verbal protocols and retrospective debriefing. *Journal of Educational Psychology*, *92*, 413–425.

Timmermans, R. E., Van Lieshout, E. C. D. M., & Verhoeven, L. (2007). Gender-related effects of contemporary math instruction for low performers on problem-solving behaviour. *Learning and Instruction*, *17*, 42–54.

Tjur, T. (2009). Coefficients of determination in logistic regression models – a new proposal: The coefficient of discrimination. *The American Statistician*, *63*, 366–372.

UNESCO (1994). *The Salamanca statement and framework for action on special needs education.* Paris: UNESCO Special Education, Division of Basic Education.

Van der Ven, S. H. G., Boom, J., Kroesbergen, E. H., & Leseman, P. P. M. (2012). Microgenetic patterns of children's multiplication learning: Confirming the overlapping waves model by latent growth modeling. *Journal of Experimental Child Psychology*, *113*, 1–19.

Van Zanten, M., Barth, F., Faarts, J., Van Gool, A., & Keijzer, R. (2009). *Kennisbasis rekenen-wiskunde voor de lerarenopleiding basisonderwijs [Arithmetic and mathematics knowledge base for primary school teacher training].* The Hague/Utrecht, The Netherlands: HBO raad/ELWIeR/Panama.

Verguts, T., & Fias, W. (2005). Interacting neighbors: A connectionist model of retrieval in single-digit multiplication. *Memory and Cognition*, *33*, 1–16.

Verschaffel, L., Luwel, K., Torbeyns, J., & Van Dooren, W. (2009). Conceptualizing, investigating, and enhancing adaptive expertise in elementary mathematics education. *European Journal of Psychology of Education*, *24*, 335–359.

Vukovic, R. K., & Lesaux, N. K. (2013). The relationship between linguistic skills and arithmetic knowledge. *Learning and Individual Differences*, *23*, 87–91.

Vygotsky, L. S. (1934/1986). *Thought and language.* Cambridge, MA: MIT Press.

Wallace, A. H., & Gurganus, S. P. (2005). Teaching for mastery of multiplication. *Teaching Children Mathematics*, *12*, 26–33.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*, 269–304.

Williams, D. M., & Jarrold, C. (2010). Predicting inner speech use amongst children with Autism Spectrum Disorder (ASD): The roles of verbal ability and cognitive profile. *Journal of Autism and Developmental Disorders*, *40*, 907–913.

Winsler, A., Fernyhough, C., & Montero, I. (Eds.). (2009). *Private speech, executive functioning, and the development of verbal self-regulation.* New York, NY: Cambridge University Press.

Wu, S. S., Meyer, M. L., Maeda, U., Salimpoor, V., Tomiyama, S., Geary, D. C., et al. (2008). Standardized assessment of strategy use and working memory in early mental arithmetic performance. *Developmental Neuropsychology*, *33*, 365–393.

Xu, R. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, *22*, 3527–3541.