MENTAL COMPUTATION FLUENCY: ASSESSING FLEXIBILITY, EFFICIENCY AND ACCURACY

Sarah Hopkins, James Russo, Ann Downton

Monash University

In this paper, we elucidate the steps taken to develop an assessment of mental-computational fluency, with the aim to capture the key features of the construct, namely procedural flexibility, efficiency, and accuracy. Over 200 children in third and fourth grade completed the assessment. Using Rasch analyses, we investigated which items contributed to producing a measure and which did not. The findings allowed us to modify the assessment and produce a unidimensional scale for measuring mental-computational fluency with addition. We discuss the value of this assessment in terms of advancing learning theory and educational practice.

PROCEDURAL FLUENCY

As in many countries, reform in mathematics education in Australia has required greater emphasis be placed on children developing proficiencies relating to understanding, fluency, problem solving, and reasoning (Australian Curriculum Assessment and Reporting Authority, 2010). This study is focused on children's development of fluency and in particular procedural fluency, defined by the National Council of Teachers of Mathematics (NCTM) as:

> the ability to apply procedures accurately, efficiently, and flexibly; to transfer procedures to different problems and contexts; to build or modify procedures from other procedures; and to recognize when one strategy or procedure is more appropriate to apply than another (NCTM, 2014, p.1).

Embedded in this definition of procedural fluency is the expectation that children will have more than one procedure for solving mathematical problems and can switch between procedures depending on problem features, and so demonstrate procedural flexibility. Children should also be able to apply an increasing understanding of mathematical ideas to modify procedures and remove redundant steps, and so increasingly rely on efficient procedures, and make few errors when doing so (displaying accuracy).

Rittle-Johnson (2017) explained how procedural fluency develops as procedural knowledge and conceptual knowledge interact and influence each other in an iterative process through experience in solving problems. The procedures and concepts exhibited by children as they build procedural fluency with mental computation are well documented in the literature, particularly for the operations of addition and subtraction, and are outlined in the next section.

2019. In M. Graven, H. Venkat, A. Essien & P. Vale (Eds.). *Proceedings of the 43rd Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 376-383). Pretoria, South Africa: PME.

Mental computational fluency

As children develop procedural fluency with mental addition and subtraction, they demonstrate a good understanding of the structure of numbers (place value), the positioning of numbers (number magnitude), and the meaning of the operations (Verschaffel, Greer, & De Corte, 2007). They apply a range of procedures to solve problems, which are often characterised in the literature as being one of three types: split strategies, jump strategies, and compensation strategies (Blöte, Klein, & Beishuizen, 2000; Torbeyns, Verschaffel, & Ghesquière, 2006). Both split and jump strategies make use of *standard* partitioning (i.e., with reference to place-value) and rearranging to make adding easier. Split strategies involve partitioning two or more addends [e.g., 34 + 48= (30 + 40) + (4 + 8)] and jump strategies involving partitioning one addend [e.g., 76 + 13 = (76 + 10) + 3]. Compensation strategies involve changing (adding to or subtracting from) one addend, to make use of a known fact, and then compensating for this change: [e.g., 17 + 9 = (17 - 1) + (9 + 1) or 17 + 9 = (17 + 10) - 1].

There is another type of partitioning strategy, often described in the literature relating to single-digit addition, referred to as decomposition strategies (Siegler, 1987). Decomposition strategies can be thought of as non-standard partitioning strategies, where at least one addend is partitioned to make use of a known fact: for example, a doubles fact [e.g., 25 + 26 = (25 + 25) + 1] or a tens fact [e.g., 47 + 6 = (47 + 3) + 3].

While there is considerable research in the field identifying the different types of strategies children use for mental computation with addition and subtraction, and the understandings children require in order to apply these strategies meaningfully, there are few suitable assessments for evaluating mental computation fluency. Some assessments depend on self-reports of strategy use to capture indicators of efficiency and flexibility (e.g., Beishuizen, 1993). These involve labour-intensive data collection methods and so are not suited to large-scale research or everyday classroom practice. Other assessments produce problematic results, including data with a bimodal distribution (e.g., Brown & Alibali, 2018), making them unsuitable for measurement purposes. More broadly, assessments designed to measure computational fluency have focused on speed and accuracy, and ignored key features of the construct, namely flexibility and efficiency (e.g., Calhoon, Emerson, Flores, & Houchins, 2007; Foster, 2018).

Baird, Andrich, Hopfenbeck, and Stobart (2017) point out the huge impact educational assessments can have on teaching and learning when they function as a communicative device, making explicit to educators what represents quality learning. These authors explicate that the notion of quality (including how proficiency of a construct is operationalised) will depend on four spheres of influence, which are theory based, empirically driven, expert devised, and policy driven. They argue that the washback effects of assessments can serve the goals of education particularly well, when learning theory and assessments (aligned with assessment theory) are strongly associated and have reciprocal effects upon each other. Relating specifically to procedural fluency, Rittle-Johnston (2017) also expressed the need for researchers to develop and validate measurement tools to build theories of mathematical development, which are more comprehensive than those that currently exist.

In the next section, we explain the steps we took to construct an assessment designed to measure mental computational fluency with addition (MCF-A). We outline findings from an analysis of responses from over 200 children to identify items that contribute to producing a scaled measure and an investigation of items that did not. It is worth mentioning that previously we pilot-tested an assessment that included both addition and subtraction items. We found children struggled with the format of the assessment and switching between operations, and so included only addition items in this version of the assessment.

ASSESSMENT DESIGN

To construct an assessment for measuring mental computational fluency with addition (MCF-A), we used a novel format for items that required students to explain a strategy used by Emma (a fictitious student). Instructions to children were:

> Emma is good at adding numbers. She uses clever strategies to make adding easier. Your job is to try and think like Emma. Explain what Emma did to get the number in the box. For example, to solve 3 + 4, Emma thinks, that is the same as [6] + 1. What numbers did Emma add together to get 6? (Answer: 3 + 3).

According to Rasch Measurement Theory (Andrich, 2016), a critical component in designing assessments to measure a construct is to generate items that range from being very easy to get right (or endorse), to items that are very difficult. Furthermore, item difficulty must be invariant regardless of who attempts these items, provided they are drawn from the population for which the assessment was designed. Items that adhere to this principle produce scores that display a Guttman structure, which is ideally needed before scores from items can be added together to form a measure. The Rasch model represents a probabilistic Guttman structure and can be compared with data from an assessment to evaluate the extent to which items on the assessment provide scores that adhere to this structure.

While items for assessing a construct on either end of a continuum are often easy to write, it is harder to write items that operate at regular intervals along the continuum. To create items for assessing mental computation fluency that function along a continuum of difficulty, we included items with single-, double- and triple- digit addends, and items with two or three addends. Since we used items with three addends, it was appropriate to include a new type of procedure, one not previously mentioned in the literature, which we labelled *noticing strategies*. Noticing strategies do not involve partitioning or changing an addend, only changing the order of addends to make adding easier. These involve combining addends that have a particular association or relationship, such as tens facts or doubles [e.g., 3 + 5 +7 = (3 + 7) + 5].

We hypothesised three factors would influence item difficulty: (i) the size of the addends, (ii) the number of addends, and (iii) the type of strategy represented.

Predicting item difficulty based of addend features was straightforward; however, previous studies have produced mixed findings in relation to which strategy is more difficult, possibly due to differences in national curricula (Verschaffel et al., 2006). We reasoned that strategies that can be used to solve both single-digit and multi-digit problems (i.e., noticing and non-standard partitioning strategies), would be easier than strategies used to solve only multi-digit problems (e.g., standard partitioning strategies), as children would have opportunities to practice these at a younger age. We used a grid (see Table 1) to help construct items that differed in difficulty. Table 1 is organised such that items differ in terms of addend features (organised in rows) and strategy type (organised in columns). The item numbers represent the predicted order of difficulty and the order of presentation in the assessment.

Noticing	Non-standard partitioning	Standard partitioning	Compensating
	1. $4+5=\boxed{8}+1$		3. $9+3=\boxed{13}-1$
	2. $4+7=\boxed{10}+1$		
4. $5+8+5=\boxed{10}+8$			
5. $6+3+6=\boxed{12}+3$			
6. $8+3+2=\boxed{10}+3$			
			7. $8+19=\boxed{28}-1$
			8. $4+28=\boxed{30}+2$
	9. $10+11=\boxed{20}+1$	10. $28+13=\boxed{38}+3$	
		11. $21+26=\boxed{40}+7$	
12. $11+3+9=\boxed{20}+3$			
13. $6+18+34=\boxed{40}+18$			
	14. $48+45=\boxed{90}+3$	17. $76+21=\boxed{96}+1$	16. $44+49=\boxed{94}-1$
	15. $36+37=\boxed{72}+1$	18. $56+33=\boxed{80}+9$	19. $22+49=\boxed{21}+50$
		20. $23+44=\boxed{64}+3$	
21. $67+45+43=\boxed{110}+45$		22. $45+67+82=\boxed{180}+14$	
	23. $235+238=\boxed{470}+3$	25. $955+445=\boxed{1000}+400$	24. $456+356=\boxed{800}+12$

Table 1: The assessment of MCF-A comprising 25 items.

METHODS

To investigate how the assessment of MCF-A functioned, we asked 203 children in Years 3 and 4 to complete it. Using convenience sampling, participants were selected from three metropolitan public primary schools in Melbourne, Australia. These schools served a range of demographics with one school community relatively advantaged, one school relatively disadvantaged, and one school community similar to the national average, based on indicators of socioeconomic status. Participation rates in each school were 53%, 33% and 71% respectively. The researchers (authors) and a research assistant administered the assessment, which children individually completed during a 60-minute mathematics class. Most children finished the assessment after 30 minutes. Data were analysed using RUMM software (Andrich, Sheridan, Lyne, & Luo, 2000; version 2030) to evaluate how well the observed data fitted expectations of the Rasch model.

RESULTS

An initial analysis of the data indicated the assessment had targeted the population well, with the exception that 12 participants did not score on the assessment. Person-ability scores ranged from 4.6 to 3.0 logits (*M* = 0.8, *SD* = 1.9) and the item-difficulty scores ranged from -2.9 to 0.4 logits (*M* = 0.0, *SD* = 1.4). The total item-trait statistic (chi-squared = 321.93, *df* = 50, *p* = 0.000) indicated some modifications to the scale were needed. Ideally, the probability statistic should be greater than 0.05 (without Bonferroni adjustment).

Noticing	Non-standard partitioning	Standard partitioning	Compensating
	1. $4+5=\boxed{8}+1$		3. $9+3=\boxed{13}-1$
5. $6+3+6=\boxed{12}+3$			
			7. $8+19=\boxed{28}-1$
			8. $4+28=\boxed{30}+2$
		11. $21+26=\boxed{40}+7$	
		17. $76+21=\boxed{96}+1$	
21. $67+45+43=\boxed{110}+45$	23. $235+238=\boxed{470}+3$		

Table 2: Misfitting items on the MCF-A

The poor functioning of item 3 and item 7, which were predicted to be the easiest for use with a compensating strategy, are noteworthy given these items were considerably more difficult than expected. Instead of being ranked 3rd and 7th in difficulty as predicted, they were ranked 15th and 18th respectively (based on the initial analysis of data, before items were removed). While it might be that recognising use of a compensation strategy was more difficult than expected, this point is not clear given these items did not fit the Rasch model. It could be that items solved using a compensating strategy are more sensitive in picking up differences in teaching approaches. In other words, a compensatory strategy may be emphasised in some classes or schools and not in others. Alternatively, it could be that these items were not well constructed.

Item response curves for Item 3 and Item 7 indicating that at least some children were guessing the answer, suggesting a problem with how the items were constructed. We reasoned that Item 3 (9 + 3 = [13] - 1) could actually be efficiently solved using a counting-on strategy (where children count on three). A better item might be 9 + 7 = [17] – 1, as a compensatory strategy is clearly more efficient for this problem than a counting-on strategy. We also noticed that Item 7 (8 + 19 = [28] – 1) might have been made more difficult (unintentionally) due to the ordering of the addends. A better item would be 19 + 8 = [28] – 1.

The inconsistent ordering of the addends might also help explain why responses to Item 8 (4 + 28 = [30] + 2) did not fit the Rasch model. The item's characteristic curve indicated it was over discriminating. This item might have distracted students of lower fluency ability and be less confusing for some students if it was written as 28 + 4 = [30] + 2 or 4 + 28 = 2 + [30].

Item characteristic curves for Items 1 and 11 indicated both items were discriminating. We reasoned that Item 1 (4 + 5 = [8] + 1) did not work well for students of higher fluency ability because some students might just know (i.e., can retrieve) the answer of nine; hence, the item was not a good example of a problem efficiently solved using a non-standard partitioning strategy. Similarly, Item 11 (21 + 26 = [40]+7) did not work well because some students might just know the answer. We examined the item character curves for the other ill-fitting items but possible explanations for misfit were less forthcoming.

After removing the nine ill-fitting items, the test-of-fit summary statistics for the modified assessment (with 16 items) produced scores that did not deviate from perfection imposed by a Guttman structure (chi-squared = 321.93, *df* = 50, *p* = 0.031). (Note, the probability statistic needed to be greater than 0.003 with Bonferroni adjustment.) All items produced scores fitting the Rach model and there was no evidence of differential item functioning according to school. As all items were dichotomous (marked either correct or incorrect), the ordering of thresholds did not have to be checked. Furthermore, a principal components analysis of residuals indicated the scale to be unidimensional. While the results are promising (illustrated in Figure 1), further work to the assessment is needed as 27 children did not score. The power of analysis of fit was good but not excellent.

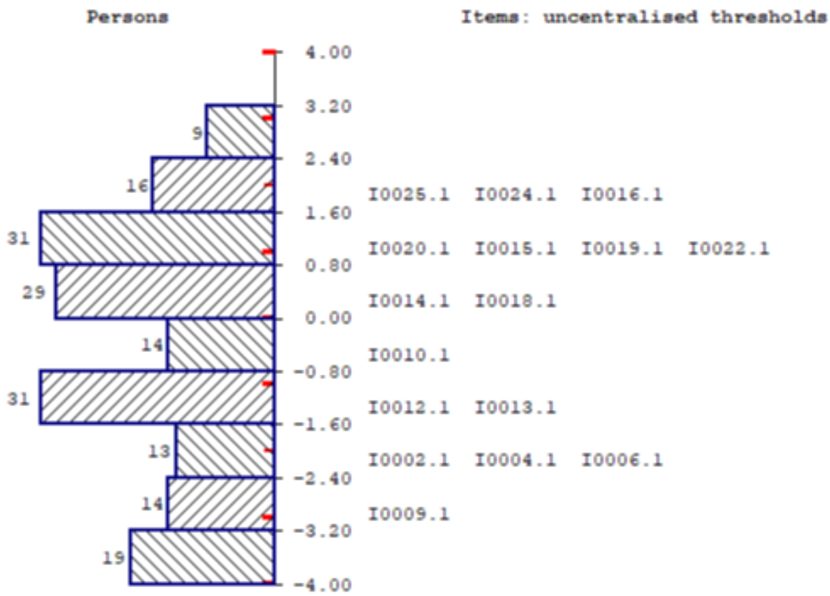

Figure 1: Item map illustrating person scores and item scores (representing item difficulty) in logits, using the modified assessment of MCF-A.

Children who found it difficult to score would have benefitted from more items that were easier and possibly from clearer instructions.

DISCUSSION

Applying Rasch measurement theory, we designed an assessment to measure computational fluency with mental addition for children in Years 3 and 4. Testing the assessment to see if it produced scores that were consistent with the Rasch model provided much useful information. The analyses identified problematic items and illuminated possible reasons for why misfit occurred. After removing these items, we found the assessment targeted the population fairly well and produced a scaled and unidimensional measure of a mental computational fluency.

We contend that this assessment of MCF-A has great potential for communicating what students need to know and be able to do. The novel format of items operationalises important aspects of procedural fluency as defined by the NCTM (2014). The MCF-A requires flexibility, as children need to generate a range of potential solution procedures when considering the strategy Emma had used. It requires knowledge of efficient procedures because the intermediate step used by Emma represents an efficient means of solving the problem. The child also has to accurately recall or calculate the appropriate number fact corresponding to the intermediate step arrived at by Emma. In this sense, the assessment addresses discipline-expert devised criteria of quality learning. It also builds on a substantial body of empirical findings revealing the different types of mental strategies children use to add together multi-digit numbers. Furthermore, we believe the assessment has much potential for improving educational practice and contributing to learning theory. It meets the criteria articulated by Baird et al. (2017) and the need expressed by Rittle-Johnston (2017) for better tools to build comprehensive, integrative theory of how fluency develops. As the removed items represented key strategy types, they were theoretically important and so need to be rewritten. The next version of the assessment is currently undergoing testing.

References

Andrich, D. (2016). Rasch rating-scale model. In W. J. Van der Linden (Ed.), *Handbook of item response theory: Volume 1, Models.* Boca Raton, Florida: CRC Press.

Andrich, D. Sheridan, B., Lyne, A., & Luo, G. (2000). *RUMM: A windows-based item analysis program employing Rasch unidimensional measurement models.* Perth: Murdoch.

Australian Curriculum, Assessment, and Reporting Authority. (2010). *The Australian Curriculum.* Retrieved January 1, 2019, from http://www.australiancurriculum.edu.au/

Baird, J.-A., Andrich, D., Hopfenbeck, T. N., & Stobart, G. (2017). Assessment and learning: fields apart? *Assessment in Education: Principles, Policy & Practice, 24*(3), 317-350. doi:10.1080/0969594X.2017.1319337

Blöte, A. W., Klein, A. S., & Beishuizen, M. (2000). Mental computation and conceptual understanding. *Learning and Instruction, 10*(3), 221-247. doi:10.1016/S0959-4752(99)00028-6

Brown, S. A., & Alibali, M. W. (2018). Promoting strategy change: Mere exposure to alternative strategies helps, but feedback can hurt. *Journal of Cognition and Development, 19*(3), 301-324. doi:10.1080/15248372.2018.1477778

Calhoon, M. B., Emerson, R. W., Flores, M., & Houchins, D. E. (2007). Computational fluency performance profile of high school students with mathematics disabilities. *Remedial and Special Education, 28*(5), 292-303. doi:10.1177/07419325070280050401

Foster, C. (2018). Developing mathematical fluency: Comparing exercises and rich tasks. *Educational Studies in Mathematics, 97*(2), 121-141. doi:10.1007/s10649-017-9788-x

National Council of Teachers of Mathematics (NCTM) (2014). *Procedural fluency in mathematics: A position of the NCTM.* Reston, VA: NCTM Ed.

Rittle-Johnson, B. (2017). Developing mathematics knowledge. *Child Development Perspectives, 11*(3), 184-190. doi:10.1111/cdep.12229

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General, 116*(3), 250-264. doi:10.1037/0096-3445.116.3.250

Torbeyns, J., Verschaffel, L., & Ghesquière, P. (2006). The development of children's adaptive expertise in the number domain 20 to 100. *Cognition and Instruction, 24(*4), 439-465. doi:10.1207/s1532690xci2404_2

Verschaffel, L., Greer, B., De Corte, E. (2007). Whole number concepts and operations. In F. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the N* (pp. 557-628). Charlotte, NC: Information Age Publishing.