The development and validation of two new assessment options for multiplicative thinking

Dianne Siemon
RMIT University
<dianne.siemon@rmit.edu.au>

Rosemary Callingham
University of Tasmania
<rosemary.callingham@utas.edu.au>

Lorraine Day
University of Notre Dame Australia
<lorraine.day@nd.edu.au>

The capacity to recognise, represent, and reason about relationships between different quantities, that is, to think multiplicatively, has long been recognised as critical to success in school mathematics in the middle years and beyond. Building on recent research that found a strong link between multiplicative thinking and algebraic, geometrical, and statistical reasoning, this paper will describe the development and validation of two new assessment options for multiplicative thinking and discuss the significance of this for the teaching and learning of mathematics in the middle years of schooling.

Introduction and Theoretical Background

Multiplicative thinking has long been recognised as a necessary foundation for fractions, rate, ratio, percentage, and proportional reasoning in the middle years (Harel & Confrey, 1994; Siegler et al, 2012; Vergnaud, 1988). However, at least 30% and up to 55% of Australian Year 8 students do not have access to this critical facility (Siemon, Banks, & Prasad, 2018). While research suggests that formative assessment can be a powerful means of improving student learning (Black & Wiliam, 1998), it would appear that this is more difficult to implement than previously thought (Smith & Gorard, 2005; Swan & Burkhardt, 2014; Wiliam. & Leahy, 2014). Hodgson et al (2014) suggests that one of the reasons for this may be that "formative assessment has been described generically rather than in subject-specific terms" (p. 168). But even where evidenced-based, subject-specific formative assessment materials have been developed, they are not necessarily taken up where schools feel pressured to prepare for high stakes assessment (Wiliam et al, 2004) or teachers lack the depth of knowledge needed to provide effective feedback (Hodgson et al, 2014).

Research-based formative assessment materials to support the development of multiplicative thinking were provided by the *Scaffolding Numeracy on the Middle Years* (SNMY) project in 2006. The materials include two validated assessment options and a Learning Assessment Framework (LAF) for multiplicative thinking that incorporates an evidenced-based learning progression and targeted teaching advice. They are appropriate for use in Years 4 to 9 and offer a valid means of identifying starting points for teaching and tracking learning over time (Siemon, Breed, Dole, Izard, & Virgona, 2006).

While the SNMY materials have been used quite widely in coaching and professional development programs, their use in secondary schools is not widespread. One of the reasons given for this is that multiplicative thinking is note seen to be relevant to the 'job to be done' (Arnett, Moesta, & Horn, 2018; Siemon, 2016; Siemon, Banks, & Prasad, 2018) even though large proportion of maths curriculum at this level is dependent upon some of multiplicative thinking (Siemon, 2013).

2021. In Y. H. Leong, B. Kaur, B. H. Choy, J. B. W. Yeo, & L. F. Wong (Eds.), Excellence in Mathematics Education: Foundations and Pathways *(Proceedings of the 43rd annual conference of the Mathematics Education Research Group of Australasia)* pp. 369-378. Singapore: MERGA.

Mathematical reasoning is another aspect of the curriculum which is not seen as a focus of mathematics teaching in middle years even though it is recognised as an important proficiency in the *Australian Curriculum: Mathematics* (Australian Curriculum Assessment & Reporting Authority [ACARA], 2016). A funding opportunity in 2014[1] afforded the possibility of investigating the development of evidenced-based learning progressions and teaching advice for algebraic, geometrical and statistical reasoning that could be seen to be more related to the curriculum and thereby more relevant to the work of secondary school mathematics teachers. This also provided an opportunity to explore the extent to which multiplicative thinking (MT) was related to mathematical reasoning (MR) by including a number of tasks from SNMY in the trials of the mathematical reasoning assessment tasks and by collecting data on both MT and MR from project schools who had not participated in the earlier RMF Priority (RMF-P) project in 2013 (Siemon, 2016).

The outcomes of the *Reframing Mathematical Futures II* (RMFII) project have been reported elsewhere (Siemon, Barkatsas, & Seah, 2019; Siemon, Callingham, Day, et al., 2018), but as the analysis of RMFII trial data suggested a strong relationship between MT and MR, a secondary analysis of these data together with combined data from the RMFII project and archived data from the original SNMY project was conducted to test the extent to which this link could be empirically established. This process resulted in the development and validation of a single, integrated scale for multiplicative reasoning that incorporated the scale for multiplicative thinking and the scales for algebraic, geometrical, and statistical reasoning (Callingham & Siemon, 2021).

At around the same time, the *Growing Mathematically – Multiplicative Thinking* (GM-MT) project was initiated by the Australian Association of Mathematic Teachers (AAMT) for the purpose of trialing a Teacher's Manual that could be used as a stand-alone guide to support the use of the SNMY formative assessment materials in secondary schools. The project team comprised the Chief Executive Officers of the AAMT (past and present), three members of the RMFII team (the authors of this paper), and a representative of ACARA. Given evidence of the strong relationship between MT and MR, it was agreed that this opportunity would be used to trial two new assessment options for MT that included MR items from the single scale for MT and MR. As a result, an application was made to amend the ethics approval in place for the ongoing data analysis work of the RMFII project to cover this aspect of the GM-MT project. The purpose of this paper is describe the processes involved in developing and validating the new options and, in doing so, to address the research question: To what extent is it possible to develop valid assessments of multiplicative thinking that incorporate aspects of algebraic, geometrical and statistical reasoning?

Research Approach

The work to be reported here was made possible by the RMF-P and RMFII projects that explored the efficacy of using the SNMY materials in secondary schools alongside the development of the evidenced-based formative assessment materials for mathematical reasoning. As indicated above, the details of this work have been reported elsewhere, however it is important to acknowledge all three projects were framed in terms of a social constructivist view of learning that acknowledges the need to identify and build on what is known (e.g. Cobb & Yackel, 1996; Shepherd & Penuel, 2018).

---

[1] The Australian Mathematics and Science Partnership Program was funded by the Australian Government Department of Education and Training from 2013 to 2017

The RMFII project used a design-based research approach (Barab & Squire, 2004; Cobb, Confrey, di Sessa, et al., 2003) involving iterative rounds of assessment and the use of Rasch modelling (Bond & Fox, 2015) to scale assessment items from easiest to most difficult for the purposes of developing, testing and refining learning progressions for mathematical reasoning (Siemon, Callingham, Day et al, 2018). A similar approach was used in the GM-MT project to evaluate the two new assessment options for MT. Interested schools were recruited through AAMT in 2019 and asked to administer and assess one of the options using the scoring rubrics provided and return the de-identified results to the project team via an excel spreadsheet. Initially 25 schools agreed to participate in this process and use the other option as a pre-test at a later date, but COVID restrictions limited the extent to which schools could contribute to the GM-MT data set and provide pre- and post-test data in 2020.

Item Selection

The two options, referred to as Option 3 and Option 4, had to be compiled such that they could be statistically linked to the existing SNMY data set for validation purposes. With these constraints in mind, the tasks (each of which comprised at least one item) were chosen from the pool of 113 validated assessment items used in the SNMY and RMFII research projects. A number of tasks from SNMY Options 1 and 2 were included to provide links among the projects. Consistent with the structure of the existing SNMY Options, an extended task and a number of shorter tasks were included in each of the new Options. As there were strong conceptual links between the SNMY and algebraic reasoning, the new extended tasks both came from the RMFII. *Trains* (Option 3) used a series of increasingly complex questions to develop generalisations about the relationships between the number of wheels and the train design. *Board Room Tables* (Option 4) considered the relationship between the number of tables in a rectangular arrangement and the number of people that could be seated. Tasks from the SNMY pool were chosen because of clear links to geometric or statistical reasoning. *Stained Glass Windows* was set in a geometric context of a triangular tessellation. *Canteen Capers* drew on the Cartesian product to identify the number of possible combinations available from a school canteen, which has links to statistical reasoning and probability. Conversely tasks from the RMFII project were chosen because of explicit use of multiplicative thinking, such as drawing names from a hat and expressing the answer as a fraction (SHAT8) and designing a package to hold a given volume of soft drink (GBEV1). All tasks, with the exception of *Skin Rash* (SRASH) and SHAT8, had multiple items. The two new Options had no overlapping items to maximise their utility as pre- and post-tests over the short period of time.

Two draft options were created (referred to as draft Option 3 and draft Option 4) and piloted in a small-scale trial for feasibility.

Pilot study

Although the numbers from the initial trial were small (n = 38; for draft Option 3 and n = 32 for draft Option 4), the Rasch analysis provided sufficient indicative information about the behaviour of both the complete draft Options and the individual items to decide whether or not they were working as intended. Each option was Rasch analysed separately to provide information about the extent to which the items worked together coherently to provide a scale. Both assessments provided good fit to the Rasch model and showed high reliability. These findings indicated that the items used were suitable for alternative assessment Options.

Table 2.
Summary Statistics for Individual Assessment Options

Option (No. of Items)	Infit (Items)	Infit zstd (Items)	Outfit (Items)	Outfit zstd (Items)	Reliability
Option 3 (17)	0.99	0.01	0.94	0.03	0.93
Persons (*n* = 38)	0.97	-0.04	0.94	0.12	0.87
Option 4 (19)	1.01	0.00	0.99	-0.03	0.90
Persons (*n* = 32)	1.04	0.11	0.99	-0.01	0.80

Note: Ideal values for Infit and Outfit are 1.00, and zSTD = 0.00. Ideal reliability coefficient = 1.00.

Overall, draft Option 4 was much harder than draft Option 3. When appropriate cut scores were applied to identify zones, this option had no items in Zone 1 and only one item in Zone 2. A revised Option 4 was developed with one of the more difficult SNMY tasks (*Tiles, Tiles, Tiles*) replaced by an easier task (*Butterfly House*).

One issue that emerged was that of the items developed for geometrical reasoning few made explicit links to MT. As a result, two new questions *Enlarging Nets* (GENLG) and *Park Map* (GMAP) were developed to address perceived gaps in the geometric aspects of MT (i.e. scale and enlargement) at an easier level than those included in RMFII. These changes were incorporated into the revised Options that were then trialed as part of the GM-MT project with students from Year 5 to Year 10 in late 2020. Tables 4 and 5 show the revised task and item selection

Table 4.
Tasks and Items for Option 3 Trial

Task	Source	Item Codes
Adventure Camp	SNMY	ADCA, ADCB
Stained Glass Windows	SNMY	SWGA, SWGB, SWGC
Relations	RMFII-Alg	AREL1, AREL2, AREL3
The Beverage Company	RMFII-Geo	GBEV1RA, GBEV1RB
Skin Rash	RMFII-Stats	SRASH
Trains	RMFII-Alg	ATRNS1, ATRNS2, ATRNS3, ATRNS4, ATRNS5, ATRNS5A, ATRNS6
Enlarging Nets	New	GENLG0, GENLG1, GENLG2, GENLG3, GENLG4

Table 5.
Tasks and Items for Option 4 Trial

Task	Source	Item Codes
Butterfly House	SNMY	BTHA, BTHB, BTHC, BTHD
Canteen Capers	SNMY	CCA, CCB
Lemonade	RMFII-Alg	ALEM1, ALEM2
Hat Chance	RMFII-Stats	SHAT8
Spy Squad	RMFII-Geo	GSPSQ7, GSPSQ8, GSPSQ9

Board Room Tables	RMFII-Alg	ABRT2, ABRT3, ABRT4, ABRT5, ABRT6, ABRT7, ABRT8
Park Map	New	GMAPA, GMAPA1, GMAPB, GMAPB1, GMAPC, GMAPC1, GMAPD

These options were trialed by the schools participating in the GM project.

Trial Analysis

As the purpose of the project was to extend the usefulness of the LAF, questions from that project were used as the anchor for the two assessment options. Because there were no common items across the two forms, a link file of student responses to the items that came originally from the LAF was created from archived SNMY data. Then all responses to Options 3 and 4 and the created link file were merged into a complete data set, so that the options were solidly linked through common items. Finally, to ensure that the existing LAF scale could be validly compared to the new scale from Options 3 and 4, an anchor file was created from the link items so that the new scale was, in effect, using the same ruler. Overall there were 4494 responses included to provide maximum data about the scale.

Rasch analysis was undertaken using Winsteps v. 4.7.1.0 (Linacre, 2020). Summary statistics for the overall scale are shown in Table 1.

Table 1.

Summary Statistics for Anchored Scale from Options 3 and 4

	Infit	Infit zstd	Outfit	Outfit zstd	Reliability
Item (*n* = 50)	1.00	-0.26	1.02	-0.07	1.00
Person (*n* =4494)	0.98	-0.03	0.97	0.09	0.82

Note: Ideal values for Infit and Outfit are 1.00, and zSTD = 0.00. Ideal reliability coefficient = 1.00.

Following this trial, all the items were behaving as expected and the revised scale was interpreted using a process of ‘segmenting the variable’ (Wilson, 1999) as reported elsewhere (e.g. Callingham & Siemon, 2021).

Although the detail is too small to be seen clearly, a small part of the Wright map produced by the software (Linacre, 2019) for all trialed items is shown in Figure 1 to provide a sense of the approach used and the relationship between the MT items (blue), the RMFII-Alg items (yellow), RMFII-Geo and new geometrical reasoning items (green), and the RMFII-Stats items. The scale on the left-hand side is in logits, the unit of Rasch analysis. Items at the bottom of the map are easy whereas those at the top are difficult. Similarly, persons located towards the bottom of the map have performed less well than persons located at the top. Where persons appear at the same logit values as an item, they have a 50% chance of achieving the score allocated to that item. the Zones are marked by horizontal boundary lines. These borders are not “hard” borders. Rather the zones provide an indication where students are in relation to the development of MT.

It is noticeable that the Geometry items are more difficult for students with no items appearing in the lower two Zones. This may be due to a lack of familiarity with geometric contexts, rather than inherent difficulty. Alternatively, the kinds of reasoning in geometry occurring in Zones 1 and 2 may rely less on numerical reasoning and more on visualisation. In these Options only two statistics reasoning items were used, although aspects of the items from the SNMY project did draw on statistical thinking.

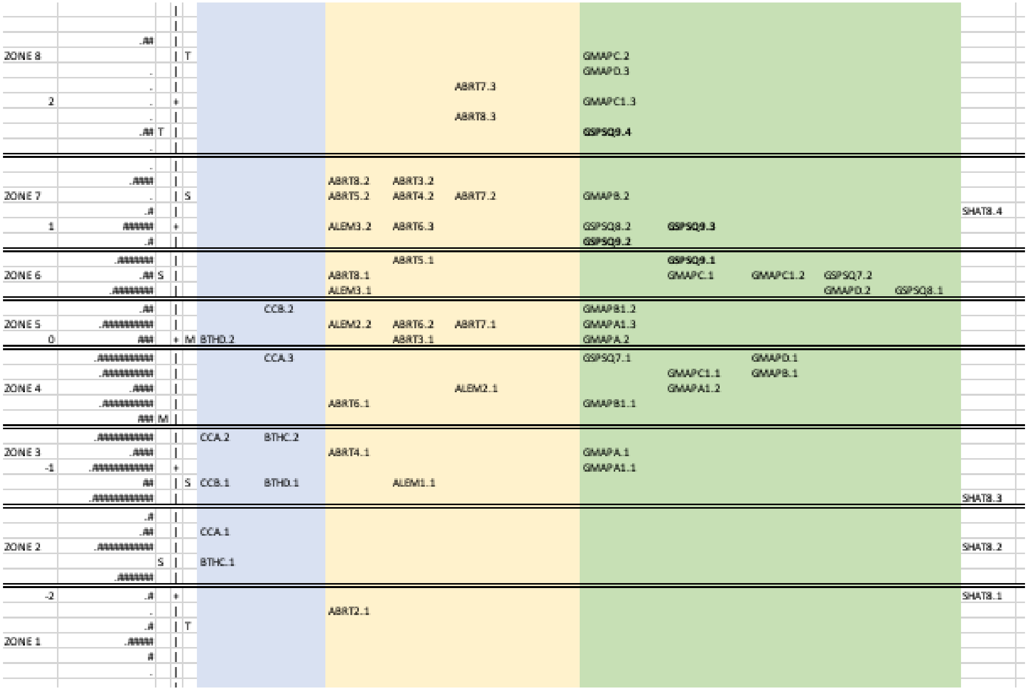

Figure 1. A portion of the Wright map from the GM-MT second trial

As part of the GM-MT project, participating schools were asked to use the assessment options as pre and post assessments to evaluate the efficacy of a targeted teaching approach to multiplicative thinking using the existing teaching advice from the Learning Assessment Framework for Multiplicative Thinking (LAF). While COVID restrictions significantly affected the number of schools who were able to provide matched data sets, the results suggest that the assessment options as trialed were working reliably and could be used to evaluate learning over time.

Discussion and Conclusion

The analysis reported in this paper has shown how assessment tasks used in previous research could be combined to create two new assessment options for multiplicative thinking that relate multiplicative thinking to algebraic, geometrical and statistical reasoning. Overall, the new scale performed in a manner remarkably similar to the existing LAF, meaning that the empirical thresholds could be retained, and the new assessment options can be used with confidence to place students within a Zone with sufficient accuracy to support targeted teaching. This is significant because secondary teachers are much more likely to see the importance of multiplicative thinking when they can visibly see its relationship to what they believe they have to teach, that is, algebra, geometry, measurement, statistics, and probability, and how this relates to mathematical reasoning more generally.

While further research and analysis is needed to test the extent to which the new options are more difficult than the existing SNMY options, this raises some questions. For instance, if it is established that they are more difficult, should the new options be ‘flagged’ as more appropriate for secondary students even though some primary school students participated in the project? As the GM-MT study was targeting the lower years of secondary schooling, is there any benefit in revising the existing SNMY options to include some of the more difficult RMFII and geometry items to better reflect the full extent to which MT is required for mathematical reasoning more generally? Should some easier reasoning type questions be developed for Year 4 to Year 6? These questions suggest there is room for more research in this area but the next step in the current process is to use the data obtained from the GM-MT trial to review and extend the original Learning Assessment Framework (LAF) for Multiplicative Thinking and to test the efficacy of using the revised framework to support a targeted teaching approach to multiplicative thinking in the middle years in a larger student population.

Acknowledgement

The authors would like to acknowledge the other members of the RMFII research team (i.e. Marj Horne, Greg Oates, Claudia Orellana, Rebecca Seah, Max Stephens, Jane Watson, Sandra Vander Pal, and Bruce White) who contributed to the development of the assessment tasks used here. The GM-MT project was conducted by AAMT with funding from the Australian Department of Education and Training. The views expressed are those of the authors, they do not necessarily reflect the views of these bodies.

References

Arnett,T., Moesta, B., & Horn, M. (2018). *The teachers's quest for progress: How school leaders can motivate instructional innovation.* Christenson Institute. https://eric.ed.gov/?id=ED593236

Australian Curriculum Assessment & Reporting Authority [ACARA]. (2016). Australian Curriculum: Mathematics F-10. https://www.australiancurriculum.edu.au/f-10-curriculum/mathematics/

Barab, S., & Squire, K. (2004). Design-based research: Putting a stake in the ground. *The Journal of Learning Sciences*, *13*(1), 1-14.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappa, 80*(2), 139–147.

Bond, T., & Fox, C. (2015), *Applying the Rasch Model: Fundamental measurement in the human sciences* (3rd ed.). Lawrence Erlbaum.

Callingham, R., & Siemon, D. (2021). Connecting multiplicative thinking to mathematical reasoning in the middle years. *Journal of Mathematical Behavior, 61*, 1-12. https://doi.org/10.1016/j.jmathb.2020.100837

Clarke, D. J. & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947-967.

Cobb, P., Confrey, J., di Sessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher, 32*(1), 9-13. https://doi/10.3102/0013189X032001009

Cobb, P., & Yackel, E. (1996). Constructivist, emergent, and sociocultural perspectives in the context of developmental research. *Educational Researcher, 31*(3), 175-190.

Griffin, P., & Callingham, R. (2006). A twenty-year study of mathematics achievement. *Journal for Research in Mathematics Education*, *37*(3), 167-186.

Harel, G., & Confrey, J. (1994). *The development of multiplicative reasoning in the learning of mathematics*. Suny Press.

Hodgson, J., Coe, R., Brown, M., & Küchemann, D. (2014). Improving students' understanding of multiplicative reasoning and algebra: Did the ICCAMS intervention work? In S. Pope (Ed.), *Proceedings of the 8th British Congress of Mathematics Education*, (pp. 167-174). https://bsrlm.org.uk/wp-content/uploads/2016/09/BCME8-Full.pdf

Linacre, J.M. (2019). *Winsteps online help manual*. Winsteps.

Shephard, L., Penuel, W., & Pellegrino, W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement Issues and Practice, 37*(1), 21-34.

Siegler R., Duncan G., Davis-Kean P., Duckworth K., Claessens A., Engel M., Susperreguy M., & Chen M. (2012). Early predictors of high school mathematics achievement. *Psychological Science, 23*(7), 691-697.

Siemon, D. (2013). Launching mathematical futures: The key role of multiplicative thinking. In S. Hernert, J. Tillyer & T. Spencer. (Eds.), *Mathematics: Launching futures*, *Proceedings of the 24th Biennial Conference of the Australian Association of Mathematics Teachers* (pp. 36-52). Adelaide: AAMT.

Siemon D. (2016). Addressing the STEM challenge though targeted teaching – What's the evidence? In *Proceedings of the 2016 ACER research conference: Improving STEM Learning- What will it take?* ACER: Melbourne.

Siemon, D. Banks, N., & Prasad, S. (2018). Multiplicative thinking a STEM foundation. In T. Barkatsas, N. Carr & G. Cooper (Eds.), *STEM education: An emerging field of inquiry*. Sense Publications.

Siemon, D., Barkatsas, T., & Seah, R. (2019). *Researching and using progressions (trajectories) in mathematics education.* Brill Sense.

Siemon, D., Breed, M., Dole, S., Izard, J., & Virgona, J. (2006). *Scaffolding Numeracy in the Middle Years – Project findings, materials and resources, Final report.* Retrieved from http://www.education.vic.gov.au/school/teachers/teachingresources/discipline/maths/assessment/Pages/scaffoldnum.aspx

Siemon, D., Callingham, R., Day, L., Horne, M., Seah, R., Stephens, M., & Watson, J. (2018). From research to practice: The case for mathematical reasoning. In J. Hunter, P. Perger, & L. Darragh (Eds.). Making waves, opening spaces. *Proceedings of the 41st annual conference of the Mathematics Education Research Group of Australasia* (pp. 40-49). Auckland: MERGA.

Siemon, D. & Virgona, J. (2002). Reflections on the Middle Years Numeracy Research Project — Is it a case of too much too soon, for too many? In B. Barton, K. Irwin, M. Pfannkuch & M. Thomas (Eds), *Mathematics education in the South Pacific. Proceedings of the 25th Annual Conference of the Mathematics Education Research Group of Australasia. (*Vol. 2, pp. 617–624). Sydney: MERGA.

Smith, E., & Gorard, S. (2005) "They don't give us our marks": The role of formative feedback in student progress. *Assessment in Education: Principles, Policy and Practice, 12*(1), 21-38.

Swan, M. & Burkhardt, H. (2014). Lesson design for formative assessment. *Educational Designer, 2*(7). https://www.educationaldesigner.org/ed/volume2/issue7/article24/

Wiliam, D., & Leahy, S. (2014). *Sustaining formative assessment with teacher learning communities.* Learning Sciences Dylan Wiliam Centre. https://www.dylanwiliamcenter.com/wp-content/uploads/sites/3/2020/10/DW02-01-Chapter-X-TLC-Paper-03-05-17-Digital.pdf

Wiliam, D., Lee, C., Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: Impact on student learning. *Assessment in Education, 11*(1), 49-65.

Vergnaud, G. (1988). Multiplicative structures. In J. Hiebert & M. Behr (Eds.), *Number concepts and operations in the middle grades* (pp. 141–161). National Council of Teachers of Mathematics.

Wilson, M. (1999). Measurement of developmental levels. In G.N. Masters, & J.P. Keeves (Eds.). *Advances in measurement in educational research and assessment.* (pp. 151-163). Pergamon.