Partial Credit in Multiple-Choice Items

Joan Burfitt
The University of Western Australia
<joan.burfitt@research.uwa.edu.au>

Multiple-choice items are used in large-scale assessments of mathematical achievement for secondary students in many countries. Research findings can be implemented to improve the quality of the items and hence increase the amount of information gathered about student learning from each item. One way to achieve this is to create items for which partial credit can be given when students select particular incorrect options. To improve the items in this way requires a critical analysis of how the items contribute to the measure of student achievement as well as extensive knowledge of the test construct.

The inclusion of multiple-choice (MC) items in assessments of mathematical understanding appears set to continue as these items are deemed to be efficient and cost-effective for the collection of evidence of student achievement according to Betts, Elder, Hartley and Trueman (2009). For the same amount of test time a broader range of content can be covered with MC items than with other types of items. However, the quality and amount of information collected about student learning can be further improved without asking more of students who respond to these types of items. Such improvement could increase the accuracy and detail of the measures of student achievement.

Multiple-choice items consist of a statement or question, known as the stem, followed by a series of numbers, words, phrases or sentences which might complete the statement or provide the answer to the question in the stem. Typically, one of these options is correct and is known as the key while the other incorrect options are called distractors. The key is generally awarded one mark and zero is allocated for all distractors or missing responses; this is described as dichotomous scoring.

Students who do not fully comprehend an item's content may have still developed some knowledge and understanding and can be considered to have partial knowledge of the concept tested in the item. When items are scored dichotomously there is no score for partial knowledge but some scoring procedures enable partial credit to be allocated.

Partial Knowledge

There are various descriptions of partial knowledge in the research literature and for Bush (2001) it involves the selection of more than one option and is identified as *liberal* MC. Candidates are allowed to select more than one answer if they are uncertain of the correct one. For four options, the candidate scores three marks if they select only the key, two marks for two options, and one mark for three options. Bush found that it took much longer to answer the questions, the instructions needed to be very clear and the weaker students did not like the format.

Partial knowledge was described by Bond et al. (2013) as the ability to eliminate some but not all of the incorrect answers. Such elimination was positively scored but the elimination of the key attracted a negative score. Bond et al. found that students preferred this form of elimination marking to the traditional single answer selection and reported that they found it less stressful and were not distracted by thinking of ideal tactics to maximise their scores.

(2017). In A. Downton, S. Livy, & J. Hall (Eds.), *40 years on: We are still learning! Proceedings of the 40th Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 117-124). Melbourne: MERGA.

The answer-until-correct approach was described by Frary (1989) as a means of rewarding partial knowledge. Students would select options until they were correct, and they scored according to the number of attempts to identify the key. At the time, this method was costly to supervise and correct and while the use of computers would make such a process more efficient, little evidence of its current use has been found in the research literature.

In this review of scoring methods, Frary (1989) reported on a method by which options were weighted and the students would score according to which option they selected. The value of an option was determined by experts but it was deemed difficult to explain the process to examinees. Briggs, Alonzo, Schwab, and Wilson (2006) studied a similar scoring process in which each option was linked to a developmental level and even though this method appeared to work for MC items in Science, it is difficult to determine how several different levels could be planned for the options for each MC item in Mathematics.

Further scoring processes which involve option weighting are described by Diedenhofen and Musch (2015). The options are weighted after the students have completed the test and the responses are analysed by examining the correlation between the frequency of option choice and total score. Diedenhofen and Musch reported increased test reliability and validity with such scoring but suggested it is not suitable for easy items.

Proportional Reasoning

Proportional reasoning has been described by Siemon, Bleckly, and Neal (2012) as a key concept for students in early secondary, "without which, students' progress in mathematics will be seriously impacted" (p. 22). It pervades all areas of mathematics for lower secondary students and underpins many aspects of the upper school curriculum including similarity, trigonometry, functional relationships and algebraic formulation. For Siemon et al. (2012), proportional reasoning "involves recognising and working with relationships within relationships (i.e., ratios) in different contexts" (p. 32).

According to Lamon (1993), students can demonstrate proportional reasoning when they understand equivalent ratios and the invariance of relationships, even though they are unable to represent the relationship using mathematical symbols. Proportional reasoning, as the ability to reason when using proportions, or to solve problems when the relationship between quantities or variables is proportional, is the definition underpinning this study. The skills and concepts that students need to develop to be able to solve a range of problems to demonstrate sound proportional reasoning include understanding and manipulating ratios, rates, fractions, decimals and percentages.

The acquisition of the skills for sound proportional reasoning occurs over considerable time and from some research studies it has been possible to identify some of the stages in this development. Students learn some aspects of a concept before being fully competent and they may be described as having partial knowledge of the concept. Such partial knowledge can be used to create distractors for MC items which can be scored with partial credit. For some of the concepts necessary to develop sound proportional reasoning there are misconceptions that are commonly held by students and these can inhibit the ability to demonstrate other skills and understandings.

Many of these misconceptions have been described in the research literature and some of these are considered in this study. Partial knowledge can be demonstrated when students recognise the need to increase or decrease a quantity but make additive errors, using addition rather than multiplication to solve proportion problems. Misailidou and Wiliams (2003) found that additive errors were the most common errors made by students aged 10 to 13 years as they solved problems on proportional reasoning. Another demonstration of partial knowledge occurs with the use of absolute rather than proportional change. Students can increase or decrease quantities but use a fixed amount rather than a proportional amount and in this type of situation would use \$15 to represent a 15% increase regardless of the starting amount. Students may also recognise an increase in size or shape but have the scale factor incorrect as when moving from linear measure to area measure.

Student development of understanding fraction operations comes in stages and a common error made by students is described by Behr, Wachsmuth, Post, and Lesh (1984) in their study where 30% of the students added the numerators and denominators to find $\frac{1}{2}+\frac{1}{3}$, giving the answer as $\frac{2}{5}$ instead of $\frac{5}{6}$. It could be concluded that another common error that students would make is to identify $\frac{2}{4}$ as double $\frac{1}{2}$.

The use of MC items to assess mathematical understanding can be improved and one way to do this is to provide a score for the partial knowledge of a concept as shown by a student when they select a distractor which shows better understanding than other distractors. Writing such distractors requires an analysis of research findings to identify ways by which students develop concepts. It is hypothesised that providing partial credit produces a more accurate measure of student ability than dichotomous scoring and allows a more efficient use of the MC items for assessment.

Methodology

To collect data for the analysis a test of sixty MC items was designed, created and implemented. Items were written using the content and proficiencies of the Western Australian curriculum in Mathematics for Years 6 to 8 (School Curriculum and Standards Authority [SCSA], 2016). The test consisted of six blocks with ten items in each block and all students were given the first block of ten items, written at the standard Year 8 level. Students were then randomly allocated to two of the other five blocks, each of which was written for a different standard, for example, Year 7 above the standard. While 860 students completed then ten items at the Year 8 standard, between 327 and 360 students completed each of the other blocks of ten items.

The items were designed to test the skills and understandings deemed necessary for the development of sound proportional reasoning which included aspects relating to decimals, percentages, fractions, proportions, ratios, rates and linear relations. For each item there were four options, the key and three distractors. One distractor was written to attract students who knew something, but not everything about the item's content and hence allowed partial knowledge to be demonstrated. This partial knowledge was deemed to be worth some credit but not as much as was awarded for the selection of the key. The author's experience as a Mathematics teacher and results of studies reported in the research literature were used to inform the creation of distractors to be awarded partial credit.

Items 9, 56, 15, and 38 are presented in Figure 1 and relate to percentages, rates, fractions and ratios respectively. For Item 9, the distractor designed to elicit greater information is Option b and students who select this option could be thinking of absolute rather than proportional change. For Item 56, it was thought that students who were not competent in adding fractions would select Option c. For the rates described in Item 15 students not recognising that the smaller floor was a quarter of the area of the larger floor, might be able to demonstrate partial knowledge by recognising that there is a factor of two in the linear measure. The distractor in Item 38, Option c, was created to allow students who are using “additive” thinking rather than using proportional reasoning to adjust a ratio, to be awarded credit for their partial knowledge.

Item 9

The **correct** answer in a student’s homework was \$744.

The question could have been

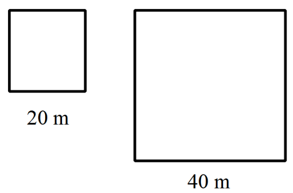
a. Increase \$600 by 24%
b. Increase \$700 by 44%
c. Decrease \$700 by \$44
d. Decrease \$800 by \$166

Item 56

Jon’s pancake recipe requires $1\frac{3}{4}$ cups of flour. How much flour will Jon need when he **doubles** the recipe?

a. $3\frac{1}{2}$ cups
b. 3 cups
c. $2\frac{6}{8}$ cups
d. $2\frac{1}{2}$ cups

Item 15



Two square gym floors need to be polished. The time estimate for the larger floor is 8 hours.
If the floors are polished at the same rate, then the time needed for the smaller floor is

a. 16 hours
b. 8 hours
c. 4 hours
d. 2 hours

Item 38

Daniel has two dogs: Benson who weighs 10 kg and Shamrock who weighs 15 kg.

Daniel gives them treats according to the ratio of their weights.

If Daniel gives Benson 12 treats, how many should he give to Shamrock?

a. 24
b. 18
c. 17
d. 12

Figure 1. Multiple-choice items from the online test.

Approval to conduct the test in Western Australian schools was obtained from the University of Western Australia (UWA), the Department of Education, Catholic Education and the Association of Independent Schools. The Year 8 students who volunteered to sit the test came from twelve different secondary schools and there were at least three schools from each sector. Given the nature of the investigation and the proposed analysis, as well as the number of students who volunteered it was not considered necessary to confirm that the sample obtained was representative of Year 8 students in the state. The online test was conducted in November 2016, a time by which the Year 8 curriculum would have been covered for most students in these schools. The survey platform which supported the creation and delivery of the online test is one licensed to UWA (Qualtrics, 2016).

The software program RUMM2030 (Andrich, Sheridan, & Luo, 2015) was used in the application of Rasch Measurement Theory to the results. Three different analyses were conducted. First, all items were scored dichotomously with one mark allocated for the selection of the key and zero otherwise. For the second analysis with polytomous scoring, two marks were allocated for the key, one mark for the distractor created to be awarded partial credit, and zero otherwise. For the third analysis, scoring was dichotomous or polytomous.

After the second analysis, an examination of the category probability curves indicated that polytomous scoring was not working for all items. These curves are provided in Figure 2 for the four items described earlier. For Items 9 and 56 there was no range of location on the continua for person ability (variable on the horizontal axis) where the probability of obtaining a score of one was higher than for all other scores. This indicates there is insufficient information to justify awarding a score of one and hence acknowledge partial credit. In these two items, the thresholds, which are the locations at which the probability is equal for adjacent scores, are said to be disordered. Evidence of ordered thresholds is seen in Figure 2 for Items 15 and 38. In Item 15 the first threshold (-1.4) is less than the second threshold (1.9). At the first threshold, the probability of scoring zero is equal to the probability of scoring one and at the second threshold, the probability of scoring one is equal to the probability of scoring two. The 35 items in which the thresholds were disordered were rescored dichotomously for the third analysis and polytomous scoring was retained for the other 25 items for which the thresholds were ordered.

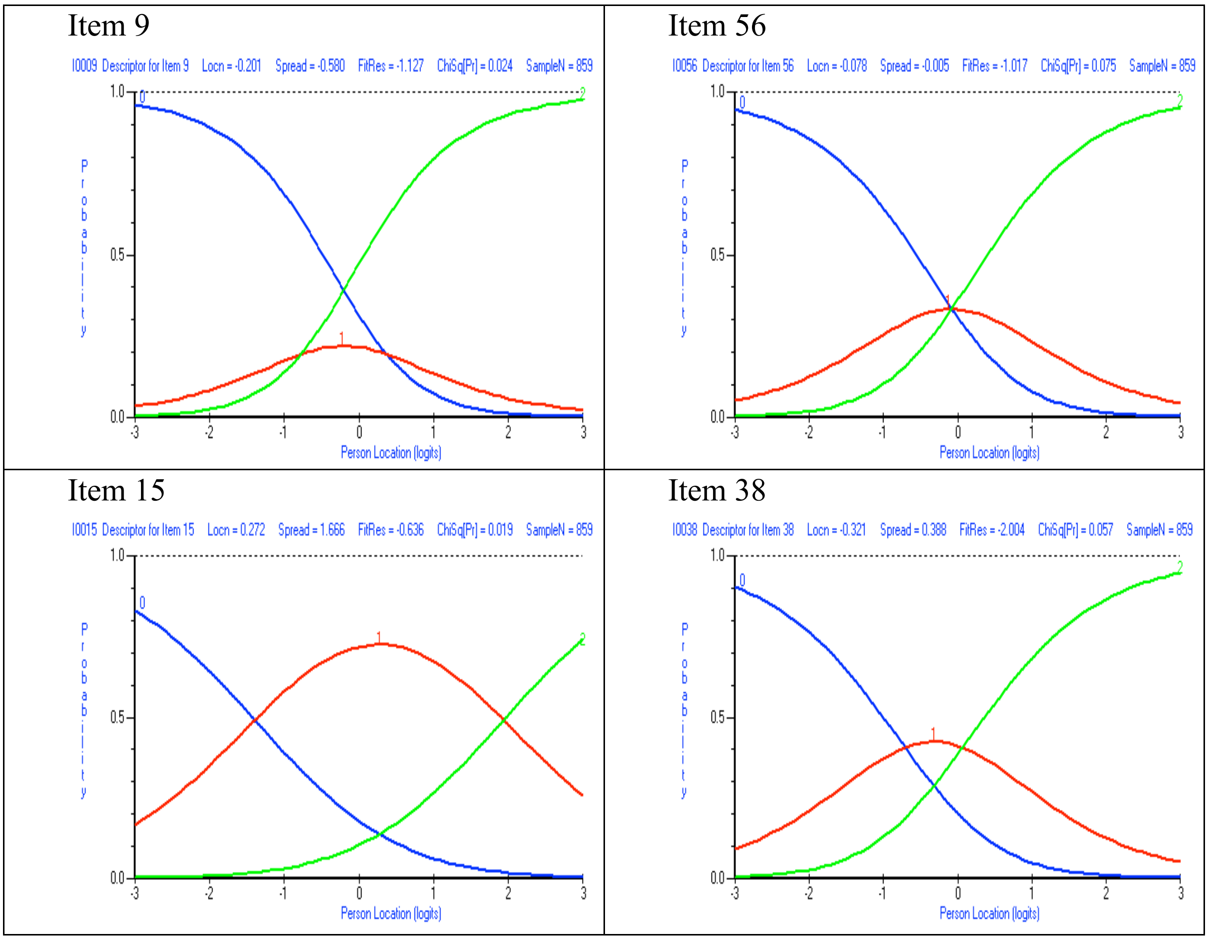


Figure 2. Category probability curves for multiple-choice items.

Results

For 25 of the 60 items, the use of polytomous scoring indicated that thresholds were ordered and this supported the proposal that students demonstrating partial knowledge could be rewarded with a score for a particular distractor other than the key. A comparison of the items where thresholds were ordered with those that were unordered has not shown any pattern that would allow a priori prediction of suitability for polytomous scoring. Items with ordered thresholds were not located in any particular area of person ability, nor concentrated in any of the particular content areas of proportional reasoning. It appears that each item needs to be analysed individually to identify the reasons why the thresholds were not ordered and why the proposed existence of partial knowledge was not confirmed.

For Item 9, the expectation that students could demonstrate partial knowledge of percentage increase with the selection of Option b was not realised. It is suggested that on the developmental pathway for most students, knowing that 44% of $700 is not $44 is learned before, or is easier than knowing that subtracting $166 from $800 is not $744. With fraction doubling in Item 56, scoring the Option c, the one designated as partial knowledge, did not appear to be justified indicating that this type of doubling is not a stage on the learning continuum.

For Item 38, where the thresholds were ordered, the selection of Option c suggests that the students may have considered that the absolute difference between the weights applied also to the difference in the ratio. Option d does not give the students the opportunity to show that they know the number of treats must increase and for Option a, the students recognise the increase but realise that the number of treats cannot be double. The type of additive thinking associated with the selection of Option c is considered as a stage in the development of proportional reasoning and for this item the award of a score for the partial knowledge is justified.

Students recognised the direction of the change in Item 15 and used the factor of 2 which was supplied for the linear measure when they selected Option c. They managed all aspects of the concept of change except for the recognition of the correct factor.

A comparison of the scales for student achievement as shown in Figure 3 indicated that awarding partial credit affected the significance of the gender differences as well as the measures of student achievement. The difference between male and female achievement was significantly higher for males with all analyses but the level of significance (p = 0.0403) was less when scoring for partial knowledge than when all items were scored dichotomously (p = 0.0105). The mean person location increased by 0.417 from -0.316 to 0.099, with the award of partial credit but the increase was greater for females than for males, 0.437 compared to 0.385.

The distribution of persons in Figure*s* 3 and 4 shows a considerable shift to the right at the lower end of the ability scale from the first to the third analysis. This result supports the expectation that the award of partial credit provides a higher level of achievement for persons of lower ability. This movement is not evident for persons of higher ability. The measurement scale appears to be more condensed with higher frequencies of person ability in the middle locations. Further evidence of the narrowing of the scale is seen in the lower standard deviation for both genders and this supports the idea that the overall variation of achievement is reduced when partial knowledge is rewarded in MC items.

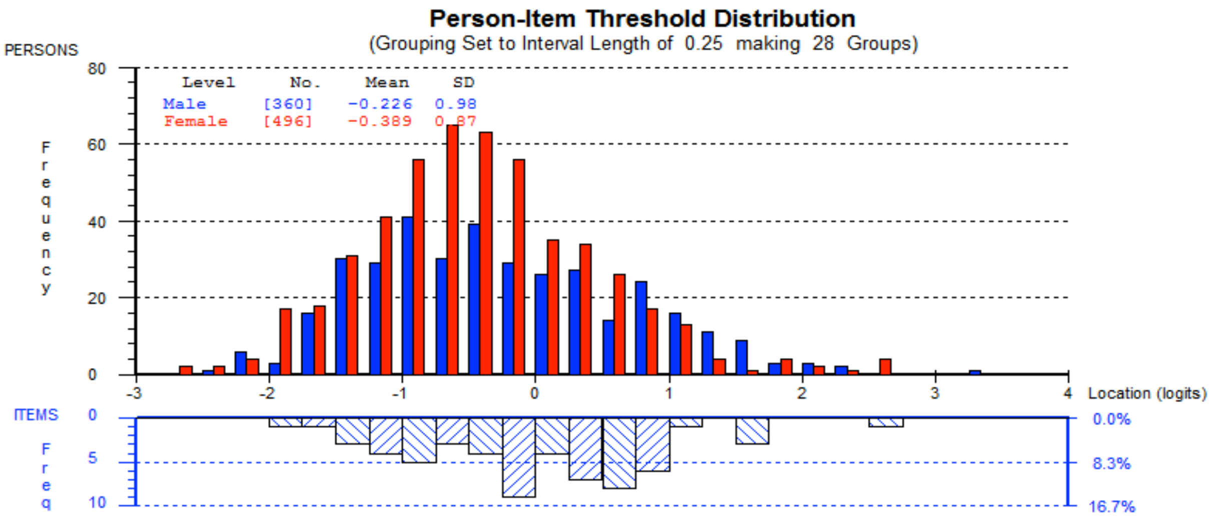


Figure 3. Person-Item distribution showing gender differences for first analysis.

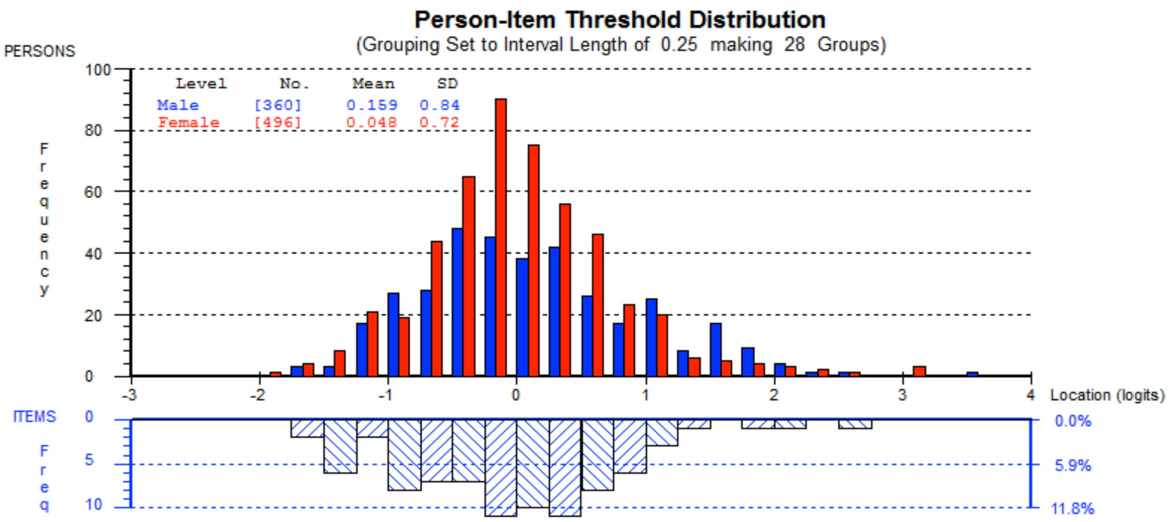


Figure 4. Person-Item distribution showing gender differences for third analysis.

Conclusion

For the allocation of credit for partial knowledge when using MC items in the assessment of mathematical understanding, there are two important considerations. First, it is desirable to have a sound awareness of the development in student understanding of the test content to be able to create items with options that reflect different levels of student ability. Second, it is necessary to critically analyse the students' responses to the item to confirm that the proposed partial knowledge represents a stage on the continuum of learning. More accurate measures of student ability can be made if credit can be given for partial knowledge and to do this with MC items allows more information about student learning to be gathered without increasing the demands on students taking the test. While greater effort is required to create such items, the time required to complete the test and the behaviour of the students in selecting the best option are not affected.

Acknowledgements

The author recognises the ongoing guidance and advice provided by Chapple Professor David Andrich and Professor Peter Merrotsy during the progress of this research. The high level of support provided by the staff and students from schools in Western Australia has resulted in a valuable source of data for this investigation. The author also wishes to acknowledge the support of the Australian Government through an "Australian Government Research Training Program Scholarship".

References

Andrich, D., Sheridan, B. S., & Luo, G. (2015). *RUMM2030: Rasch unidimensional models for measurement.* Perth: RUMM Laboratory. Retrieved from www.rummlab.com

Behr, M. J., Wachsmuth, I., Post, R., & Lesh, R. (1984). Order and equivalence of rational numbers: A clinical teaching experiment. *Journal for Research in Mathematics Education, 15*(5), 323-341.

Betts, L. R., Elder, T. J., Hartley, J., & Trueman, M. (2009). Does correction for guessing reduce students' performance in multiple-choice examinations? Yes? No? Sometimes? *Assessment and Evaluation in Higher Education, 34*(1), 1-15. doi:10.1080/02602930701773091

Bond, A. E., Bodger, O., Skibinski, D. O. F., Jones, D. H., Restall, C. J., Dudley, E., & van Keulen, G. (2013). Negatively-marked MCQ assessments that reward partial knowledge do not introduce gender bias yet increase student performance and satisfaction and reduce anxiety. *PLoS ONE, 8*(2), e55956. doi:10.1371/journal.pone.0055956

Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic assessment with ordered multiple-choice items. *Educational Assessment, 11*(1), 33-63. doi:10.1207/s15326977ea1101_2

Bush, M. (2001). A multiple-choice test that rewards partial knowledge. *Journal of Further and Higher Education, 25*(2), 157-163. doi:10.1080/03098770120050828

Diedenhofen, B., & Musch, J. (2015). Empirical option weights improve the validity of a multiple-choice knowledge test. *European Journal of Psychological Assessment*. doi:10.1027/1015-5759/a000295

Frary, R. (1989). Partial-credit scoring methods for multiple-choice tests. *Applied Measurement in Education, 2*(1), 79-96.

Lamon, S. J. (1993). Ratio and proportion: Connecting content and children's thinking. *Journal for Research in Mathematics Education, 24*(1), 41-61.

Misailidou, C., & Williams, J. (2003). Diagnostic assessment of children's proportional reasoning. *Journal of Mathematical Behaviour, 22*, 335-368. doi:10.1016/S0732-3123(03)00025-7

Qualtrics (2016). *Qualtrics insight platform* [computer software]. Provo, UT: Qualtrics Labs. Retrieved from http://www.qualtrics.com

School Curriculum and Standards Authority. (2016). *Mathematics curriculum.* Retrieved from http://k10outline.scsa.wa.edu.au/home/p-10-curriculum/curriculum-browser/mathematics-v8

Siemon, D, Bleckly, J., & Neal, D. (2012). Working with the big ideas in number and the Australian curriculum: Mathematics. In B. Atweh, M. Goos, R. Jorgensen, & D. Siemon (Eds.), *Engaging the Australian curriculum mathematics: Perspectives from the field*. Retrieved from http://www.merga.net.au/