Cognitive interviews for reviewing multiple-choice items in mathematics

Joan Burfitt
The University of Western Australia

During the development of a test of multiple-choice items, cognitive interviews were conducted with students from lower secondary school. The purpose of these interviews was to confirm that the prospective respondents' interpretation of mathematics test items was consistent with the interpretation intended by the item writer. The conversations with the students provided unexpected and intriguing insights into the way students think and speak about mathematical concepts: ideas worthy of comment and further research. The benefits and limitations of using cognitive interviews to improve assessment items, and the recommended procedures for conducting such interviews, had been identified from earlier research studies. With these findings, a framework was developed for designing and implementing a series of interviews with ten Year 9 students in Western Australia. The items reviewed by these students were intended for an online test of the skills and understandings essential for the development of sound proportional reasoning for Year 8 students. The responses to the interview questions not only provided useful suggestions on how to improve the test items and on the suitability of the test instrument, but they also provided valuable information about the students' understanding and communication of what they had learned.

Introduction

Multiple-choice items predominate in local, national and international assessments of mathematics for Western Australian students. These assessments include the *Online Literacy and Numeracy Assessment* (School Curriculum and Standards Authority, 2017), the *National Assessment Program – Literacy and Numeracy* (National Assessment Program, n.d.-a), and the *Trends in International Mathematics and Science Study* (Martin, Mullis & Foy, 2013). To gain credit for their response to a multiple-choice item, students need to select the option which correctly completes the sentence or answers the question posed in the introduction. The focus of this study into the function of multiple-choice items was to investigate ways of increasing the quality of information that can be gathered when students respond to this item format. One aspect of this study involved designing and constructing multiple-choice items for which partial credit can be awarded when a student selects an incorrect option (distractor) which indicates that they have some, but not all the knowledge (partial knowledge) needed to identify the key (correct response).

To confirm that it is possible to provide partial credit for multiple-choice items, an online test of 60 items was designed and created. Each multiple-choice item had four options: the key, one distractor that was specifically designed to attract students with partial knowledge, and two other distractors. Proportional reasoning was chosen as the context for investigating the function of multiple-choice items because it is an important concept for students to develop in early secondary school. Without good understanding of proportional reasoning, student progress in mathematics in upper secondary school is limited. Proportional reasoning is defined as the ability to recognise and 'work with relationships between relationships' (Siemon, Bleckly & Neal, 2012, p. 25) and for Year 8 students the associated skills, as described in the curriculum (Australian Curriculum, Assessment and Reporting Authority, n.d.), relate to calculations with fractions, decimals, percentages, linear relationships, ratios and rates. Different types of partial knowledge that can be demonstrated when solving problems involving proportional reasoning were identified as: the use of additive rather than multiplicative strategies; the consideration of the numbers as being absolute rather than proportional; the recognition of correct calculations but the inability to execute them; and, the ability to estimate but not determine an exact value (Burfitt, 2017).

Before items are used for collecting data, they should be reviewed and checked for clarity and for their relevance to the construct (Peterson, Peterson & Powell, 2017). There is evidence that multiple-choice items for mathematics assessments are trialled during item development (National Assessment Program, n.d.-b) and that students are given items to complete during trialling programs. However, reports of students responding to questions about their interpretation of multiple-choice items in mathematics have not been located. For this study on improving the function of multiple-choice items there were three independent reviews. In two reviews, experienced educators provided feedback on the items' content and on their suitability for the target audience. For the third review, cognitive interviewing, students were asked a series of questions to investigate their interpretation of the items. In this third review, the subject of this article, the students were involved in examining the items and in providing feedback on the items before the final data collection.

Cognitive interviewing, or cognitive testing (Willis & Artino, 2013), refers to a set of techniques used in conversations with survey respondents to interrogate their understanding of the items before the final survey is implemented (Tourangeau, Rips & Rasinski, 2000). Even though the final items were used to collect data from the Year 8 students, the Year 9 students were chosen to review the items because the interviews were conducted some months before the end of the school year to allow time for editing and the Year 8 students had not covered the test content. Furthermore, the purpose of these interviews was to identify any unforeseen ways by which the students might have interpreted the test items and not to test the students' knowledge of the item content. The findings were used to review and edit the items before the final survey was conducted.

Cognitive interviews

The techniques defining cognitive interviews have been described by Tourangeau et al. (2000) and they include probing and thinking protocols which are designed to discover how respondents interact with survey questions. These interactions involve four different cognitive operations, namely, comprehension, recall, judgement and response. Conducting the interviews involves focusing on understanding the respondent's interpretation of the language used, on appreciating the role of the respondent's memory, and on gaining insight into their perception and judgement as they answer the interviewer's questions.

Karabenick et al. (2007) identified six cognitive aspects of the information processing model used by respondents:

1. read the item and retrieve relevant words from memory;
2. interpret the item and store this in working memory;
3. retrieve further relevant information from stored memory;
4. read, interpret and store in memory the options provided in the multiple-choice item;
5. simultaneously hold all the gathered information in working memory; and
6. evaluate the stored information and choose a response.

Describing interviews as cognitive is an attempt to conceptualise the processing that occurs when the person interviewed is presented with an item and then provides a response. This manner of interviewing, only recently referred to as *cognitive,* was extensively used by Piaget (1952) in conversations with subjects while studying the development of children's thinking. For Piaget the focus of such interviews was to locate evidence of the underlying processes used by children to solve problems. However, in item development the focus is on the cognitive processes involved in drawing out evidence of children's thinking (Karabenick et al., 2007). The use of cognitive interviews in the development of survey items has become more systematic and widespread since the convening of the seminar on the *Cognitive Aspects of Survey Methodology* (CASM) in 1983-4 (Campanelli, 1997; Desimone & Le Floch, 2004; Willis, 2015). Collins (2003) recommended that this provision of further information on the mental processes used by individuals as they respond to questions is a valuable diagnostic tool for pre-testing surveys and questionnaires, and should be viewed as a standard component of developing such instruments.

Purpose and benefits of using cognitive interviews

The use of cognitive interviews to provide a structured and systematic approach for identifying how respondents will interpret items is reported in several studies of item development (Collins, 2003; Fortus & Vedder-Weiss, 2014; Gehlbach & Brinkworth, 2011; Gerber & Wellens, 1997; Gray, Blake & Campanelli, 2014; Koskey, Karabenick, Woolley, Bonney & Dever, 2010; Wildy & Clarke, 2009; Willis, 2015; Wininger, Adkins, Inman & Roberts, 2014). Conducting interviews with individuals or with small focus groups before the items are used to collect data for analysis enables the researcher to review, evaluate and improve the test instrument. The evidence gathered during these interviews can be used to check that the respondents interpret the items as intended by the writer, that this interpretation is consistent between respondents, and that the respondents have all the information needed to determine an answer (Peterson et al., 2017).

Potential problems with item interpretation include over-challenging vocabulary and ambiguity of context or meaning (Gehlbach & Brinkworth, 2011). Psycholinguistic features including sentences with numerous logical operators, too many adjectives or adverbs, verbs used as nouns, complex sentence structure, and the presence of words that are infrequently used, can cause difficulties with comprehension and slow the respondent's processing time (Lenzner, Kaczmirek & Lenzner, 2010). To make items realistic for students, familiar contexts are often used; however, this can be problematic for students if they focus on the context rather than the mathematics (Greenlees, 2010). The extent to which respondents interpret items as intended by their authors has been described as a measure of cognitive validity (Karabenick et al., 2007; Koskey et al., 2010; Wildy & Clarke, 2009). Editing items to remove problems identified during the interview can increase the cognitive validity and provide users with greater confidence in the accuracy of the measurement scale.

Conducting cognitive interviews provides an opportunity to identify conceptual problems with the construct and to provide information that could be used to improve an item, as well as to optimise validity before a costly data collection (Karabenick et al., 2007). Testing the sensitivity of respondents to particular items, assessing item format and identifying an improved order of item presentation are further benefits of conducting these interviews (Gerber & Wellens, 1997). The cognitive interview processes also facilitate the identification of items likely to induce a response error, as well as indicate the variation of item interpretation among respondents (Wildy & Clarke, 2009). It is the first step in determining the face validity of an item before data are collected during piloting to undertake preliminary psychometric validation.

If only piloting is used for item review, the information gathered does not indicate the range of interpretations for an item, nor the level of confidence with which the participants respond (Wildy & Clarke, 2009). Furthermore, information to explain potential misfit to the measurement model is limited. The cognitive interview process is deemed to be essential as it provides greater insight into misleading instructions and item interpretation than piloting or trialling (Desimone & Le Floch, 2004). It supports the development of better items and measuring instruments as well as an increased understanding of measurement error (Blair & Conrad, 2011).

Limitations of using cognitive interviews

Several factors can limit the quality of cognitive interviews and thus affect the accuracy and reliability of the information collected. Seeking individuals' awareness of what they are thinking is described as metacognition and is "difficult for most people" (Anderson & Thomas, 2014, p. 9) because they are not often asked to explain or elaborate their thinking. Händel and Dresel (2018) reported that low-performing mathematics undergraduates had greater difficulty in recognising that they had the knowledge to respond to an item than the high-achieving students. One would expect that school-aged students would experience the same difficulty. Young participants may also not feel at ease when conversing with adults and may provide limited responses or even be reluctant to volunteer their time. This can affect the size and diversity of the sample population and the quality of the data gathered. If the interview questions relate to content not covered in the recent past, the student may not remember all aspects of the concepts covered in the item. Each interview can take considerable time for the researcher and requires the accurate recording and analysis of the conversation. For some interviews this might generate additional and significant costs.

Collins (2003) described cognitive interviews as limited because the interaction is essentially verbal, but not all cognitive processes can be verbalised. Furthermore, verbalisation can be difficult for some students and it may be affected by the student's mood, their environment and the previous questions posed (Gerber & Wellens, 1997). The quality of the data collected during the interview process may be diminished by two respondent behaviours, namely, satisficing and acquiescence. Satisficing occurs when the respondent does not use the minimum processing power necessary to generate an appropriate response and is satisfied to simply provide an answer without trying to address the question (Gray et al., 2014). When a respondent cannot determine the correct answer, they may acquiesce by choosing a "neutral" response, for example, by selecting an option in a particular position or by agreeing to a statement without too much thought. These actions can affect the quality of the respondent's answers and hence reduce the accuracy of the information.

Empirical evidence to indicate the number of interviews needed is limited. Blair and Conrad (2011) suggested that while a small sample may be sufficient, the size also depends on the nature of the items and the ability of the respondents to detect problems with the items. Interviewing can continue with sample populations until no more errors are identified but this is unlikely to be cost-effective. Peterson et al. (2017) described cognitive interviewing as only one step in the item review process and hence agreed with the current practice of interviewing between five and fifteen respondents.

There are few reports on the verification of the outcomes of cognitive interviewing and on the analysis of the information received. Problems with item interpretation are generally not quantifiable and the analysis of the data generated tends to be subjective. The information gathered is specific to the nature of the items and relevant information can be lost if codes are used to classify the participants' responses (Ryan, Gannon-Slater & Culbertson, 2012). Revised items are unlikely to be subject to further analysis, so evidence of improvement may not be available. As the conduct of cognitive interviews in the review of test items is still a relatively new process, the techniques are not very standardised. Willis (2015) claimed that despite these limitations the use of cognitive interviewing has greatly enhanced our understanding of the causes of measurement error and has helped researchers to improve items. Hence, the process is still considered to be a very worthwhile activity.

Designing cognitive interviews

The two techniques commonly used during cognitive interviews are described as *think aloud* and *probing* (Tourangeau et al., 2000). When thinking aloud, respondents are asked to report all that comes to mind. The contributions are free-flowing, and the role of the interviewer is to listen and to record but not to comment. Koskey et al. (2010) reported the tendency for students to go off-track when responding freely, talking about their personal stories and related examples rather than the question posed. This tendency to be diverted can cause the interviewer to lose control of the conversation and to gather much irrelevant material which may make the data analysis problematic.

In the second of the commonly-used techniques, probing, the respondents are asked specific questions about the items. These verbal probes can be used to clarify the meaning of terms, to investigate the consistency with which terms are understood, to explain why a particular answer is chosen, and to describe the processes by which an answer is sought (Fortus & Vedder-Weiss, 2014). Probes can provoke an awareness of issues unlikely to surface during a free-flowing think aloud process. Verbal probing is considered easier for the respondent because it is directed to a confined area of knowledge. Planning is essential for interviews consisting of verbal probes and the interviewer needs to be conversant with the items, aware of potential respondent behaviour, and able to control the progress of the interview (Willis, 2015). Probes can be scripted whereby all respondents are asked the same questions or they can be conditional on the response to an earlier probe. Maintaining the flexibility of being able to create a new probe in response to a student's answer can be an advantage when the student's response to a question indicates that further elaboration would support item improvement.

Various types of probes and question formats have been described in the research literature and some examples, adapted from those of Collins (2003, p. 235) and Willis (2015, p. 37), are provided in Table 1. Knowing the type of probe is not essential for designing an interview but it provides guidance for the development of the interview questions. It is important to use probes which stimulate the students' thinking of what they understand by the item, which extract this information during the interview, and which identify those item features that cause variation in the way the item is interpreted.

Table 1: Types and examples of probes

Type	Examples
General	How easy or difficult do you find this question to answer? Why do you say that?
Paraphrase	What is this question asking? The respondent may be asked to use their own words to restate the item or to say what the item is about.
Meaning	What to you is a "proportion"?
Elaborative	Why do you say that?
Process	How did you determine your answer?
Retrieval	How did you remember that? How did you calculate your answer?
Evaluative	Do you feel this question is easy or not easy to answer?
Confidence judgement	How sure of your answer are you?
Hypothetical	What type of diagram would be useful for this item?
Recall	What types of triangles were you thinking about?

Adapted from Collins (2003, p. 235) and Willis (2015, p. 37)

Interview questions may involve asking the students to read the item aloud, to say if they agree or disagree with the statements provided, to describe what this item is asking of them, and to name which answer they would choose (Campanelli, Gray, Blake & Hope, 2016; Gehlbach & Brinkworth, 2011; Karabenick et al., 2007). Conditional probes may be invoked in response to the respondent's first answer, for example, if they select the correct option in a multiple-choice item, then they may be asked to justify their selection.

The framework recommended by Cyr, Dion, McDuff and Trotier-Sylvain (2012) for the design and planning of the interview was adapted for this study. The process contains a sequence of activities by the interviewer: they present themselves to the respondents; they review the rules of the interview with the respondents; they build rapport with the respondent by asking an easy practice-run question; they ask the open-ended questions and then the specific probing questions; they bring the conversation to neutral ground before terminating the interview. These guidelines were followed but the use of the more open-ended questions was scheduled to occur after the specific probing to prevent students going off-track too early in the interview.

Method

For the study of multiple-choice items and how they function in the assessment of students' understanding of proportional reasoning, a test of 60 items was created and the items were reviewed before being used with Year 8 students in Western Australia. As part of the review process, a cognitive interview was designed to collect information from secondary students about the suitability of the multiple-choice items. The interviews were conducted with Year 9 students who volunteered to participate in the study and the information gathered from the students as they responded to questions about the items, was used to edit and improve the items.

The multiple-choice items in the online test

For the creation and presentation of the multiple-choice items for the test, the guidelines proposed and tested by Haladyna, Downing and Rodriguez (2002) were applied. The recommendations included placing the main idea in the stem (introductory text) rather than in the options, keeping choices independent, writing stems and options using positive language, and ensuring all incorrect answers were plausible. All items had four options from which the students were to choose the one correct answer. For each item, one of the incorrect options was written so that it was "partially correct". An example of one of the items used in the cognitive interviews is provided below.

ITEM 5.3
Phil is packing up to move. He has two boxes which are both rectangular prisms. He estimates the larger box is twice as high, three times as long and twice as wide. Approximately how many times greater is the capacity of the larger box?

(a) 24
(b) 12
(c) 14
(d) 7

Figure 1: Sample item from student test

For this item the key is (b), and students who do not appreciate the multiplicative nature of scaling would have selected option (d) and partial knowledge recognised because they have used each scaling factor additively. In the planned study the students selecting (b) would score 2 marks, students selecting (d) would score 1 mark and those selecting (a) or (c) would not receive any score.

The interview questions

The interview questions were adopted or adapted from those used in studies by Collins (2003, p. 235), Campanelli (1997, p. 21), Karabenick et al. (2007, p. 143), and Willis (2015, p. 37). The suggestions by Wininger, Adkins, Inman and Roberts (2014) that (a) students are asked how they think an item could be changed, and (b) students are invited to make any additional comments, were also included. The interview questions were designed to check that the wording of the item was clear, to collect suggestions for improving the test, and to gauge students' opinions about the perceived level of difficulty of the test for the Year 8 students. There were 22 questions in the interview and most of these required a short response about an individual item; only two questions referred to the whole test. All students were given the same set of questions which can be seen in Table 2, where both questions and results are shown. Starting questions were relatively easy and only required a simple response, while the later questions involved more probing and requests for explanations. The questions and the order in which they were presented were designed to put the students at ease and to keep the interviews on track. Where the students only needed to agree or disagree with the statement, they were specifically reminded each time that they only needed to provide a one-word answer.

Interview respondents

Year 9 students were chosen for the cognitive interviews because the interviews were scheduled for the second last term of the school year and these students would have been taught the curriculum content covered in the test. The Year 8 students would have only covered about three-quarters of the year's work and they would have probably been taught different aspects relating to proportional reasoning. By interviewing Year 9 rather than Year 8 students, there was less risk that any small sample of respondents contained a significant number of students yet to develop their knowledge of the Year 8 curriculum content for proportional reasoning. Volunteers were sought from a secondary school in Western Australia where the school staff had agreed to support the interview process. As the students were volunteers they were likely to be confident and competent mathematicians who were interested in supporting this research. A sample size of 15 was proposed and this would have enabled each item to be examined by five different students.

Conducting the interviews

The procedures for collecting information from students complied with the national requirements for ethical conduct as outlined in the National Statement (National Health -

Table 2: Interview questions and quantification of students' responses (N=10)

Questions		Achieved	
No.	Text	Yes	No
1	Please will you read this item out loud?	10	0
2	Please will you say this item again, but this time use your own words?	10	0
3	Do you understand what the wording of this item means?	8	2
4	Tell me what you think this item is about.	10	0
5	Do you know what this item requires you to do?	10	0
6	Tell me what you think this item requires you to do.	9	1
7	Can you eliminate one of the options?	10	0
8	Can you tell me which option you already know is incorrect?	10	0
9	Is this item an easy one to understand?	8	2
10	Is it easy to work out the correct option for this item?	9	1
11	Is there anything you would like to see changed in this item?	10	0
12	Is this item easier than the previous one?	6	4
13	Describe what could be done to make this item easier for you.	8	2
14	The word "ratio" is used in this question. Do you know this term?	10	0
15	As you read this item tell me the first thought that comes into your mind.	9	1
16	As you read this item, try to determine the correct response. Tell me what you are thinking as you do so. (can explain thinking)	10	0
17	Which option would you choose? No need to explain why. (chooses correct option)	2	8
18	Select an answer. Please can you explain why you chose that particular option? (can justify choice of option)	8	2
19	Please explain to me how the first option differs from the last one?	8	2
20	What further knowledge do you need to work out the correct response for this item?	3	7
21	Do you think this test is too hard for Year 8 students? Explain.	5	5
22	In what ways do you think this test can be improved before it is given to Year 8 students later in the year?	5	5

and Medical Research Council, Australian Research Council, & Australian Vice Chancellors' Committee, 2015). Approval to conduct the research on behalf of The University of Western Australia was granted by Human Ethics from the University's Office of Research Enterprise. The Director of Catholic Education, Western Australia gave consent to approach the principals of Catholic schools to request their participation in the data collection. All letters of communication to the school, teachers, parents and students were approved by the university. Written permission was obtained from each child participating in the interviews, their parents and the school principal.

The interviews lasted between 20 and 25 minutes and were held over a period of two weeks. The students could elect to be interviewed on their own or with another student, but all chose individual interviews. The interviews were conducted during the school day and took place in a breakout room situated in the school library. The students' responses were not coded or digitally recorded but handwritten by the interviewer who was also the principal researcher. While there were enough volunteers for the planned 15 interviews, the final number of interviews conducted (N=10) was affected by the lack of signed parental permission slips needed to satisfy the ethics requirements of the University and the school system. Instead of each item being reviewed by five different students, the number of different inspections varied from two to five. All students stayed for the duration of the interview and no student took up the option to bypass questions that they did not wish to answer. All appeared very willing to provide their support for the item review.

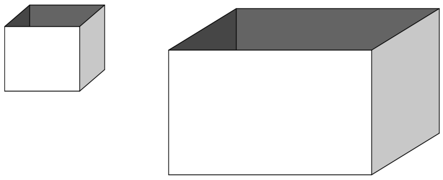
After the initial introductions the students were informed that the purpose of the research was to improve the way multiple-choice items are written and that the aim of the interviews was to check that the students understood the items in the way the writer had intended. The students were reminded that it was not an aim of the interview to test mathematical knowledge and skills. The guidelines for the conduct of the interview were presented to each student and a written copy left in view of the student for the duration of the interview. As the online test progressed the items appeared on the computer screen, one at a time and the student was asked the next question from the list shown in Table 2. At the end of the questions and after thanking the students for their support for the research, a question was posed, or a comment made to return the conversation to neutral ground.

Findings

The results of the cognitive interviews provided in Tables 2 and 3 indicate that only minor revisions to the items were needed. For the items presented, all students were able to read and paraphrase them, describe their content, and eliminate and identify incorrect options. All students claimed that they knew the mathematical terms highlighted, could determine the action required to select a correct response, and stated that the item under review did not need changing. Furthermore, most of the students said that they could understand the wording; they could describe what they thought the item required; they could describe what could be done to make the item easier; they could provide ideas mathematically related to the item; they could provide a sensible explanation for choosing an option; and they could explain the difference between two options. Most of the students claimed that the item under consideration was an easy one to understand and that there was no problem in selecting the correct response to the item.

Given the recommendations made by the students, diagrams were added, sentences reworded and language simplified where the meaning of the item was not clear. Some sentences were shortened and some repositioned with increased spacing. Item 5.3 was initially presented to the students as shown in Figure 1 but as one student suggested a diagram would assist, the item was edited, and the final version is shown in Figure 2. The sentences were simplified, and a diagram of the boxes was provided as a stimulus to consider the effect of scaling on capacity. As most students showed a good understanding of the items which they reviewed and indicated that they thought that the test was at the appropriate level for Year 8 students, the basic mathematical content was unaltered in response to conducting these interviews.

ITEM 5.3



Phil has two boxes which are both rectangular prisms.
The larger box is twice as high, three times as wide and twice as long as the smaller box.
How many times greater is the capacity of the larger box?
(a) 24
(b) 12
(c) 14
(d) 7

Figure 2: Edited version of Item 5.3

The test items had been designed and written to provide familiar contexts and there was no evidence that these contexts would have caused any mathematical difficulties. However, in one item where the calculation of the distance travelled was required, one student was puzzled because the person named in the item cycled to a venue and then walked back. The student queried the reason for this. The item was not edited because attempts to explain this difference may have caused further confusion. Most students would have been able to justify the change in mode of travel.

Evidence of good thinking was present and well expressed in some of the students' explanations, and one could get a sense of what the student was thinking even though the explanation was limited. Some of the good mathematical thinking can be seen in the two students' quotes provided below.

Example 1: Asked to think about comparing five-twelfths and three-eighths one student said they were thinking "about which was closer; the 5 to the 12 or the 3 to the 8".

Example 2: When identifying the radius of the circle one student commented that it was "20 cm because 20 is a third of 60 and the small circle's radius is a third of the large circle's radius".

In the item reproduced below two students gave limited explanations of the differences between the first and last options.

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

Student 1: "Last option wrong, starts at 700, take away rather than add. But first one starts at 800, take away 100, goes straight to 700. Also wrong minus 156 not 56".

Student 2: "Top one starts higher and decreases by a higher number. Bottom one starts lower and decreases by a lower number".

Insight into the students' mathematical thinking expressed during the interviews could be grouped into three distinct categories. First, there was a focus on looking for algorithms to assist with the calculation required to identify a correct response. Second, the mathematical language used to express answers to interview questions was inaccurate and suggested a lack of understanding. Finally, the students in their conversations and selections of responses demonstrated what had been earlier defined as partial knowledge of proportional reasoning.

Items requiring calculations for speed and percentages were those prompting students to search for algorithms and where students indicated that they needed a formula they were unable to recall the correct rule. The responses provided in Table 3 were mostly to the interview questions 15, 16 and 18, where students were asked to justify their choice of options or to describe what they were thinking as they read the item or tried to answer it. The explanatory language used was often brief, simplistic and inaccurate, and at times it was difficult to determine the meaning intended by the student.

Table 3: Students' explanations of their thinking

Looking for algorithms	Use of inaccurate language
thinking about the formula. not know straight away	divide 8 cups into each quarter
little triangle DTS (average speed)	the hours and the rate for it to be polished
try and find the % or 1% and the increase is 100+	trying to divide to figure it out – want to know his in g, number kg = 2.5, 1 kg = ? Sorry can't explain
with percentages comes fractions so I'd be using fractions to represent percentages then I'd change the fractions back to percentages	multiplication one because times is used to simplify it down
a formula to minus 8% from $60	make two fractions the same and figure it out that way, make the denominators the same.
use percentage change	not representing the same as last hours not the same as 3 times 6
you have to memorise lots of formulas	convert the fractions so they are the same

Support for further investigation into rewarding partial knowledge was evident in the selection of correct and incorrect options. Only two students identified a correct option, and of those who were incorrect, most chose the option designated as partial knowledge. One aspect of partial knowledge in proportional reasoning relates to the use of additive thinking in multiplicative situations and this was evident in one student's response. The cost was given for 8 kg of fruit and the student was asked to identify the cost for 10 kg. Rather than thinking of the need to determine the cost per kilogram, the student's focus was on finding the cost of the two extra kilograms. For another item students were given a diagram of two concentric circles and the fraction showing that the radius of the small circle divided by the radius of the larger circle was one third. They were told that the radius of the larger circle was 60 cm and asked to select the radius of the small circle from 10, 20, 30 and 40 cm. When asked to identify the correct option and explain their selection one student responded "10 cm, no, 30 cm, small circle is 1:3 and 3 x 30 = 60". The thinking of this student is correct, and a calculation error is preventing access to the correct answer. This thinking was classified as the type of partial knowledge where the calculation is recognised but inaccurately performed.

Conclusions, limitations and implications

The use of cognitive interviews in this study was successful. The students responded well to the interview process and they either showed strong affirmation for the items in their current format, or they provided valuable information to revise the items. It is possible, however, that students have simply agreed with a statement when they could not form an opinion (that is, they acquiesced), and that they did not give their full attention to an item presented and the question asked (that is, they satisficed). The extent to which these actions occurred was not determinable and this issue is noted for future study. For the sample of interviewees to represent the students responding to the final version of the items, the cognitive interviews should be conducted at the same time as the final test, but this is impractical when items need to be edited. Furthermore, additional details of student characteristics would need to be collected for both the interviews and the test data collection, for example, gender, socio-economic status and language facility. While valuable information was provided by the students and then used to improve the test items, other observations and discoveries cannot be assumed to be representative of students in Year 8 and Year 9, but they are important indicators of students' thinking and warrant further research and consideration for teaching.

Many students indicated partial knowledge of concepts when asked to explain their thinking and when selecting incorrect options which contained some, but not all aspects of the correct response. The items had been created to recognise partial knowledge and these findings support the claim that items can be created for this purpose. When multiple-choice items that identify and reward this partial knowledge are used in tests of mathematical understanding, more information can be gathered about student learning without increasing the length of the test.

The unexpected benefits of this study relate to the incidental discoveries about student metacognition and the provision of ideas for future research. Knowing more about the relationship between student ability to verbalise mathematical processes and student achievement could provide valuable information to improve teaching and learning. The extent to which student explanations were affected by their capacity with literacy and their familiarity with the English language was not considered for this study and warrants further investigation. Another topic for future investigation is the alignment of the development of conceptual understanding with the strategic recall and use of algorithms.

The interview responses indicate that investigating the approaches that students take when selecting options in multiple-choice items could be beneficial. When asked what they were thinking in trying to determine the correct response, one student replied "elimination. Go to the first one and see if it's appropriate for the way I figure it out". Research into students' thoughts and actions for the selection of a correct response could provide evidence for planning how to educate students about responding to multiple-choice items in assessments of mathematics. The following information could be realised from further research in this area: identifying the approach that students take when selecting an option in multiple choice items, knowing what students are thinking as they read an item, knowing whether students determine the answer and then locate it among the options offered, determining how students distinguish between reasonable and improbable answers, and knowing to what extent the students use elimination of answers they know to be incorrect. Cognitive interviewing would be ideal for the collection of such information.

Acknowledgements

The author acknowledges the assistance of the Australian Government through an "Australian Government Research Training Program Scholarship". The author is grateful for the support received from Chapple Professor David Andrich and Professor Peter Merrotsy, as well as from the staff and students who organised and participated in the cognitive interviews.

References

Anderson, D. & Thomas, G. P. (2014). 'Prospecting for metacognition' in a science museum: A metaphor reflecting hermeneutic inquiry. *Issues in Educational Research,* 24, 1-20. http://www.iier.org.au/iier24/anderson.html

Australian Curriculum, Assessment and Reporting Authority. (n.d.). *Australian curriculum: Mathematics*. https://www.australiancurriculum.edu.au/f-10-curriculum/mathematics/

Blair, J. & Conrad, F. G. (2011). Sample size for cognitive interview pretesting. *Public Opinion Quarterly,* 75(4), 636-658. http://dx.doi.org/10.1093/poq/nfr035

Burfitt, J. (2017). Partial knowledge of understandings needed for proportional reasoning. *Proceedings of the British Society for Research into Learning Mathematics,* 37(2). http://www.bsrlm.org.uk/wp-content/uploads/2017/08/BSRLM-CP-37-2-06.pdf

Campanelli, P. (1997). Testing survey questions: New directions in cognitive interviewing. *Bulletin of Sociological Methodology,* 55(1), 5-17. https://doi.org/10.1177/075910639705500103

Campanelli, P., Gray, M., Blake, M. & Hope, S. (2016). Cognitive interviewing as tool for enhancing the accuracy of the interpretation of quantitative findings. *Quality and Quantity,* 50(3), 1021-1040. https://doi.org/10.1007/s11135-015-0188-y

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research,* 12(3), 229-238. http://dx.doi.org/10.1023/A:1023254226592

Cyr, M., Dion, J., McDuff, P. & Trotier-Sylvain, K. (2012). Transfer of skills in the context of non-suggestive investigative interviews: Impact of structured interview protocol and feedback. *Applied Cognitive Psychology,* 2694)*,* 516-524. https://doi.org/10.1002/acp.2822

Desimone, L. M. & Le Floch, K. C. (2004). Are we asking the right questions? Using cognitive interviews to improve surveys in education research. *Education Evaluation and Policy Analysis,* 26(1), 1-22. https://doi.org/10.3102%2F01623737026001001

Fortus, D. & Vedder-Weiss, D. (2014). Measuring students' continuing motivation for science learning. *Journal of Research in Science Teaching,* 51(4), 497-522. https://doi.org/10.1002/tea.21136

Gehlbach, H. & Brinkworth, M. E. (2011). Measure twice, cut down error: A process for enhancing the validity of survey scales. *Review of General Psychology,* 15(4), 380-387. https://doi.org/10.1037/a0025704

Gerber, E. R. & Wellens, T. R. (1997). Perspectives on pretesting: "Cognition" in the cognitive interview? *Bulletin of Sociological Methodology,* 55(1), 18-39. https://doi.org/10.1177%2F075910639705500104

Gray, M., Blake, M. & Campanelli, P. (2014). The use of cognitive interviewing methods to evaluate mode effects in survey questions. *Field Methods,* 26(2), 156-171. https://doi.org/10.1177/1525822X13492703

Greenlees, J. (2010). The terminology of mathematics assessment. In L. Sparrow, B. Kissane & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 218-224). Fremantle: MERGA. https://www2.merga.net.au/documents/MERGA33_Greenlees.pdf

Haladyna, T. M., Downing, S. M. & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education,* 15(3), 309-333. https://doi.org/10.1207/S15324818AME1503_5

Händel, M. & Dresel, M. (2018). Confidence performance judgement accuracy: The unskilled and unaware effect revisited. *Metacognition and Learning,* 13(3)*,* 265-285. https://doi.org/10.1007/s11409-018-9185-6

Karabenick, S. A., Woolley, M. E., Friedel, J. M., Ammon, B. V., Blazevski, J., Bonney, C. R., de Groot, E., Gilbert, M. C., Musu, L., Kempler, T. M. & Kelly, K. L. (2007). Cognitive processing of self-report items in educational research: Do they think what we mean? *Educational Psychologist,* 42(3), 139-151. https://doi.org/10.1080/00461520701416231

Koskey, K. L. K., Karabenick, S. A., Woolley, M. E., Bonney, C. R. & Dever, B. V. (2010). Cognitive validity of students' self-reports of classroom mastery goal structure: What students are thinking and why it matters. *Contemporary Educational Psychology,* 35(4), 254-263. https://doi.org/10.1016/j.cedpsych.2010.05.004

Lenzner, T., Kaczmirek, L. & Lenzner, A. (2010). Cognitive burden of survey questions and response times: A psycholinguistic experiment. *Applied Cognitive Psychology,* 24(7)*,* 1003-1020. https://doi.org/10.1002/acp.1602

Martin, M. O., Mullis, I. V. S. & Foy, P. (2013). TIMMS 2015 Assessment design. In I. V. S. Mullis & M. O. Martin (Eds.), *TIMMS 2015 Assessment Frameworks* (pp. 85-100). http://timssandpirls.bc.edu/timss2015/downloads/T15_Frameworks_Full_Book.pdf

National Assessment Program (n.d.-a). Numeracy. http://www.nap.edu.au/naplan/numeracy

National Assessment Program (n.d.-b). *Schools involvement in the National Assessment Program. Fact sheet.* http://www.nap.edu.au/docs/default-source/resources/fact_sheet_nap_small_scale_tests_file_alt.pdf

National Health and Medical Research Council, Australian Research Council & Australian Vice-Chancellors' Committee (2015). *National statement on ethical conduct in human research*. National Health and Medical Research Council. https://nhmrc.gov.au/about-us/publications/national-statement-ethical-conduct-human-research

Peterson, C. H., Peterson, N. A. & Powell, K. G. (2017). Cognitive interviewing for item development: Validity evidence based on content and response processes. *Measurement and Evaluation in Counseling and Development,* 50(4)*,* 217-223. https://doi.org/10.1080/07481756.2017.1339564

Piaget, J. (1952). *The child's conception of number*. London: Routledge & Kegan Paul.

Ryan, K., Gannon-Slater, N. & Culbertson, M. J. (2012). Improving survey methods with cognitive interviews in small- and medium-scale evaluations. *American Journal of Evaluation,* 33(3), 414-430. https://doi.org/10.1177/1098214012441499

School Curriculum and Standards Authority (WA) (2017). OLNA [The online literacy and numeracy assessment]. https://senior-secondary.scsa.wa.edu.au/assessment/olna

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*. Online publication: Mathematics Education Research Group of Australasia. https://www2.merga.net.au/sites/default/files/editor/books/1/Chapter%202%20Siemon.pdf

Tourangeau, R., Rips, L. J. & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511819322

Wildy, H. & Clarke, S. (2009). Using cognitive interviews to pilot an international survey of principal preparation: A Western Australian perspective. *Educational Assessment, Evaluation and Accountability,* 21(2), 105-117. https://doi.org/10.1007/s11092-009-9073-3

Willis, G. (2015). *Analysis of the cognitive interview in questionnaire design*. Oxford University Press. https://global.oup.com/academic/product/analysis-of-the-cognitive-interview-in-questionnaire-design-9780199957750?cc=au&lang=en&

Wininger, S. R., Adkins, O., Inman, T. F. & Roberts, J. (2014). Development of student interest in mathematics scale for gifted and talented programming identification. *Journal of Advanced Academics,* 25(4), 403-421. http://dx.doi.org/10.1177/1932202X14549354

Willis, G. B. & Artino, A. R. (2013). What do our respondents think we're asking? Using cognitive interviewing to improve medical education surveys. *Journal of Graduate Medical Education,* 5(3)*,* 353-356. https://doi.org/10.4300/JGME-D-13-00154.1

Joan Burfitt is enrolled as a PhD candidate at The University of Western Australia. Her interests include the teaching, learning and assessment of mathematics for students in the middle years. Joan's current research focuses on improving the function of multiple choice items in assessing student achievement.
Email: joan.burfitt@research.uwa.edu.au

Please cite as: Burfitt, J. (2019). Cognitive interviews for reviewing multiple-choice items in mathematics. *Issues in Educational Research*, 29(2), 346-362.
http://www.iier.org.au/iier29/burfitt.pdf