IMPROVING LEARNING THROUGH INTERACTIVE MULTIPLE CHOICE QUESTIONS WITH CONFIDENCE MEASUREMENT

Graham Farrell, Vivienne Farrell
Swinburne University of Technology
Melbourne, Australia
gfarrell@swin.edu.au; vfarrell@swin.edu.au

ABSTRACT
Multiple Choice Questions (MCQs) offer benefits to both students and educators, however they have been criticized for not offering the ability to test partial knowledge, having limited ability to assess deeper levels of understanding and by nature reward guessing. This research initially discusses and analyses various alternative online MCQ assessment scoring regimes, specifically implemented to eliminate the negative aspects of MCQs. It then compares these previous initiatives to an innovative MCQ with Confidence Measurement (MCQCM), designed and implemented for tertiary level Computer Science students, with the objective of this meta-cognitive approach to assessment is to provide richer formative feedback as a means to improve student learning and enhance the educational experience.

KEY WORDS
MCQ, confidence measurement, assessment, meta-cognitive, computer science education.

1. Introduction
It is the ongoing challenge of educators to develop, trial and implement innovative educational approaches that facilitate the changing demands of the learners, learning institutions and the general society. The advent of the use of technology in the delivery and assessment of education has necessitated research into the evolving concerns and the development of best practices.
Assessment plays a critical role in the educational process as both a means of grading (summative) and supplying valuable feedback (formative). Technology has increased the expectations for affective assessment systems to be available, encouraging self-assessment at all stages of the learning experience as an integral part of the learning activities rather than an interruption [1]. Enabling a continuum of self-assessment and feedback throughout learning has been identified as the most influential contributor to the student’s progress [2], highlighting areas of understanding, misunderstanding and complete ignorance, underpinning rather than undermining the student’s confidence [3].
MCQs are frequently used in traditional education forums for both formative and summative assessment, as they can assess broad fields of learning in a compact system while being quick to assess with inherent objectivity and provide good feedback to the students [4].

Traditional assessment can often fall short of their goals, as demonstrated by the use of the Multiple Choice Question (MCQ). MCQs are criticized for encouraging students to direct their learning to studying the lower levels of Blooms Taxonomy of Educational Objectives [5] such as rote learning formulas and terminologies, failing to prepare for the questions based on case studies at the higher Blooms levels of “application and synthesis”. The MCQ format is also considered to permit, and even encourage guessing as part of the testing strategy, as the actual learning is reliant on the student being informed of the correct answer to the question. Some educators employ the “Correction for Guessing” formula, designed to counteract the “Noise” generated from guessing [6]. However, correction for guessing does not enable demonstration of partial knowledge and does very little to encourage students to report their true levels of knowledge and their confidence in their response.

1.1 Confidence Measurement for MCQs
The merit of incorporating confidence measurement with consequential penalties for wrong answers in MCQs has come under discussion spurred by the introduction of technology simplifying the ability to capture and display the relevant data [7], [8], [9], [10]. Criticism of this approach includes encouraging students to only train to achieve the highest grade [11] and not improve their knowledge. Diamond and Forrestor [12] define knowledge as asking the question “What do you know?” followed by the meta-question “How sure are you of the answer to the question about what you know?” as they consider the registration of confidence to be a significant indication of a true level of knowledge. Davidoff [13] considers medical knowledge to be often “incomplete, ambiguous and conflicting” and therefore the standard MCQ testing method does not facilitate or reflect the students’ level of knowledge. Farrell [14] states that encouraging guessing can often lead to over confidence in incorrect knowledge, with miss-calibration being equally as concerning as lack of knowledge.
Further, many instructors are becoming aware of the benefits of shifting the responsibilities of learning into the student’s hands [15]. There is also recent recognition of
“Meta-cognitive” learning, which concentrates on “thinking about their own thinking” [16]. This includes “self-awareness, self-inquiry, self-monitoring and self-regulation, “which “promotes high levels of commitment, persistence and involvement in learning”, fostering self-appraisal and self-regulated learning”.

With the concerns of educators in mind there have been substantial efforts to overcome the short fallings of the MCQs while harnessing the positive attributes they offer through the use of technology. This paper presents and discusses some of the previous innovative work in this field to assist in informing educators and those interested in pursuing further critical analysis of assessment options using confidence measurement.

2. Previous Work on Innovative Assessment

The following examples vary in structure, as some are based on MCQs with four options, others three options and in one case based on true or false answers. The discussion of the scoring technique is in context to the design of the application.

2.1 MCQs Designed to Eliminate Guessing

MCQ scoring has long been discussed in relation to models to eliminate guessing, employing complex scoring algorithms producing a more precise reflection of the student’s understanding of the underlying concepts. MCQs traditionally require the student to identify the correct response from a selection of answers with the score based on the criteria of being correct or incorrect.

Pollard [17], Hobson [18], Bush and Frandsen et al. [19] all produced alternative MCQ scoring techniques designed to minimise random guessing based on a reward and penalty structured scoring system.

In particular Pollard [17] achieved this by allocating a positive score for each correct answer and a negative score for each incorrect answer. As Pollard states: “As guessing has nothing to do with knowledge, it makes sense to design a paper that will minimise the effects of guessing.”. Pollard’s approach to assessment produced a series of penalties that were proportionally less than the rewards for correct responses. Consequently, it only depleted, not negated, the overall score if a correct option was identified.

Educational institutions, such as the Australian Mathematical Association National Mathematics Quiz, have used Pollard’s scoring to eliminate guessing. Pollard’s system relies on the combination of boxes ticked, encouraging identifying both correct and incorrect answers, producing many variations that have to be considered. A complex algorithm, as demonstrated by Table 1, calculates the resulting grade. To explain Pollard’s scoring approach we first must consider all of the possible responses to a question with 4 options. Table 1 demonstrates the possible ordered responses (Resp 1 to 10) for a question from the candidate, a simple example where the last option D (in green) is correct and the other options A, B, C (in red) are incorrect. Pollard relies on a complicated process of allocation of partial scores for a student’s correct identification of both correct and incorrect answers, facilitated by assigning $k_i$ values to the various responses, culminating in 9 $k_i$ values, where $k_i$ to $k_9$ contribute to the positive marks given and $k_i$ to $k_9$ contribute a negative effect to the score. A value is assigned to a constant $k$, used as the basis of the formula to calculate the final grade. The $k$’s (all positive) displayed in the last row indicate the formula for the score given for the response directly above.

After consideration to the expected score, E(S), for an individual who has no knowledge and is guessing Pollard derives the following sets of equations to be optimised in order that any candidate has no expected gain from randomly guessing. These equations must all be less than or equal to 0, which requires the assigned values of $k_i$ to $k_9$ (those that contribute a negative affect to the score) be selected to ensure this to occur, hence:

$$3k_1-k_2 \leq 0; \ 2k_3-k_5 \leq 0; \ k_2-k_4 \leq 0; \ k_1+k_2-k_3-k_7 \leq 0; \ k_2+k_2-k_8 \leq 0; \ k_3-k_9 \leq 0;$$

Pollard produces a number of solution sets satisfying these constraints, identifying two preferred sets that provide the most effective results, basing his final decision on the need to recognise partial knowledge and maximization of the minimal score for incorrect responses. The final calculated scores are displayed in column 1 of Table 1, with elimination of any positive gains for guessing.

Pollard’s work paved the way for other scoring systems to be developed by establishing a mathematically valid technique of using penalties in an attempt to deter students from guessing, while recognising partial knowledge. However, the scoring is difficult to apply, confusing for the instructor, and most importantly, more so for the student, possibly distracting from their attention to the question at hand.

However, there have been concerns raised with the level of complexity in the scoring calculations, as the student must fully understand the consequences of his/her action and the possible outcome when doing a test.
2.2 Enhanced Assessment Incorporating Confidence Measurement

The following discussion introduces some of the activities of the pioneers of assessment with confidence measurement. Brown et al. [20] produced an MCQ calibrated scoring system “that encouraged honesty”, designed to permit the student to register their level of confidence in choosing an answer, which severely penalised the participant if they registered high confidence in an incorrect answer and equally rewarded high confidence in a correct choice.

Paul [10] used an interactive Computer Based Alternative Assessment (CBAA) system of scoring for an MCQ format with three options and only a single correct answer, where the student nominated a position on lines joining the apexes in a triangular shape. Each apex represented three optional answers, A, B or C, as seen in Figure 1a, where A is the correct option.

![Figure 1a: CBAA Triangle answers options at each apex.](image1)

![Figure 1b: CBAA Triangle showing strength of belief P(A) that A is correct associated with each region.](image2)

The lines joining the apexes are proportionally divided into segments, creating 16 areas, requiring the student to nominate their level of confidence by placing the cursor within a grid area aimed to reflect the confidence of their choice, as the closer the student positions the cursor to the vertex the more confident they are that it is the correct answer. The corresponding probabilities for each zone when the correct answer is A are demonstrated in Figure 1b. The top apex where the correct option is placed has the probability of 1 while the other apexes are assigned zero. There are various other probabilities assigned depending on the proximity to the correct apex, notably placing the cursor in the middle, declaring that the student is not sure of the correct option, assigned .33, as expected. A student placing the cursor in any of the 16 zones creates the three-element vector < PA, PB, PC >, where PA, PB, PC is the probability registered reactive to A, B and C respectively. Consequently, positioning the cursor in the middle of the triangle gives the vector <.33, .33, .33>. The scores generated are based on a logarithmic function that calculates a set of non-linear results. As an example a student that positions the cursor close to the correct answer A at the top vertex, say .8 in Figure 1b, receives the score of 92/100, which reflects their strong commitment to a correct answer. The same positioning towards an incorrect answer, .2 or 0 gives a score of 46/100 and 0/100 respectively, depending on whether the cursor is positioned favouring the correct answer or the other incorrect answer.

Paul relies heavily on the notion that a scoring system should be “admissible” or “proper” as defined by probability theory, stating that a student who exhibits high belief in the likelihood of a correct answer should be rewarded higher than those who “shade” their reporting with lower levels of confidence. Paul’s adopted scoring system was developed to quantify uncertainty into numerical probabilities for representation of intelligence. For example, a student (1), who indicates 40% for an incorrect answer, is not completely wrong nor is it right. They are better than another student (2) who indicated 80% for the same wrong answer, and should accordingly be graded higher; however, they are not as good as the third student (3) that registered 10% for the same wrong answer, who receives a higher mark than student (1).

Paul’s scoring is based on assigning credit according to a scheme similar to wagering, where there are various wages with various odds. Paul, leverage greatly off Shannon’s Information Theory preferring to assign the function $f(u) = 1$ based on logarithmic scoring yielding through integration the logarithmic scoring system. Authors can provide an in-depth explanation of the mathematics that underpin this scoring but for this research the derived important equations only shall be presented.

$$p_i \quad n \quad p_j$$

Payoff if $i^{th}$ event occurs = $\int f(u) \ du / u - \sum_{j=1}^{n} \int f(u) \ du / 1/n$

yielding through integration the logarithmic scoring system. The expected adjusted payoff is: $n$

$$\text{Expected Profit} = \log n - (\sum_{j=1}^{n} p_j \log p_j)$$
(Payoff on the \(\ell^{th}\) event, with \(n\) alternatives and \(p_\ell\) is the probability of the \(\ell^{th}\) alternative).

This final interpretation of the formula for scoring is adapted by Paul to produce the following three equations for scoring the CBAA, where \(p_i\) is the probability ascribed to for alternative \(x\), \(n\) is a normalisation constant and \(k\) is a range constant.

Score if A is correct \(= n + k \log_2 (3 p_A)\)
Score if B is correct \(= n + k \log_2 (3 p_B)\)
Score if C is correct \(= n + k \log_2 (3 p_C)\)

Paul claims that the CBAA package offers discrimination between finer grained states of knowledge, greater disclosure of student’s ability to apply their knowledge and increased awareness of the students own knowledge state. Importantly, he believes that using the traditional 0-1 scoring system loses the ability to discriminate between states of knowledge. Paul uses common expressions to support the student through the experience, such as, “I strongly believe B to be correct”, “I believe C to be correct but I can’t distinguish between A and B” or “From what I know each alternative seems equally likely to be correct.” Each expression is equated to equivalent probabilities. In his diagnosis of previously adopted scoring systems Paul justifiably identifies concerns of the traditional “Number Right” grading system, where the same mark is assigned to those who have knowledge and those that have guessed, or alternatively grouping together those with complete misinformation, those with some misinformation and those who guess incorrectly.

The more recent, extensive work of Gardner-Medwin [8] has had significant influence on the direction of this research. Gardner-Medwin’s uses a scoring system that is “proper”, rewarding a student for demonstrating their true beliefs and being truly honest. He argues that a scoring system should use incentives to encourage the participants to expose their real state of knowledge. In contrast there is a double negative penalty score for high confidence in an incorrect answer. He considers delusion to be worse than misconception. Gardner-Medwin states that “lucky guesses are not the same as knowledge and that confident, wrong answers need special attention”, claiming guessing can have extreme detrimental effects on the student’s learning. He maintains that educators describe the degrees of belief that a student has about a true statement as either they have Knowledge, Uncertainty, Ignorance, Misconception or Delusion. Gardner-Medwin assigns probabilities for the truth to the above student states that they range from 1 to 0, where \(p=1\) is knowledge, \(p=0.5\) is acknowledged ignorance and \(p=0\) is delusion, with uncertainty placed between .5 and 1 and Misconception between 0 and .5. Delusion is of extreme concern as it is a total belief in something that is false. In light of this, ignorance is not the worst state to be in. Misconceptions (\(p=.33\)), having a level of confidence in an incorrect answer can be an obstacle in learning, especially when attempting to build high levels of learning. It is for this reason that he introduces a safe zone for students who are not confident of their knowledge on a particular area by creating a non-penalising area for low confidence in an incorrect answer.

His scoring technique has three options, for simplicity of use, permitting the participant to register 3 distinct levels of confidence, high, moderate or low, required to be registered after answering a True/False or traditional MCQ question of a clusters of 3 or 4 possible answers. Gardner-Medwin’s scoring system has some distinguishing features, as seen in Table 2 and Figure 2, in that it rewards the student, who selects the correct answer proportionally to the confidence registered, that is a grade of 3 for high confidence registered as C=3, a grade of 2 for moderate confidence registered as C=2 and a grade of 1 for low confidence registered as C=1. Importantly in contrast, it severely penalises the student who registers a high level of confidence (C=3) for an incorrect answer with a score of -6, moderately penalises the student who registers moderate confidence (C=2) for an incorrect answer with a score of -2 and does not penalise the student who registers low confidence (C=1) for an incorrect answer by giving them 0, previously referred to as the safe zone. In summary Gardner-Medwin’s scoring reward for a correct choice stays proportional for all of the options (3, 2, 1) while the penalty score does not (0,-2,-6), as seen in Figure 2. Gardner-Medwin’s work is considered by some to be highly controversial, as he applies some server penalties. He justifies his choice by stating that his scoring is “properly motivating” and that “lucky guesses, are not the same as knowledge”. The graph in Figure 2 demonstrates the optimal path for a student to maximize their score. Gardner-Medwin references the work of Shannon’s theory of information [21], investigating the relationship between the scores and the appropriate information-theoretic measure of lack of knowledge, demonstrating a pleasing relationship of the subjective probability assigned to the correct truth value for a proposition.

Table 2:
Gardner-Medwin’s 3 levels of confidence.

<table>
<thead>
<tr>
<th>UCL Confidence-based scoring scheme</th>
<th>Confidence Level</th>
<th>Score if Correct</th>
<th>Score if incorrect</th>
<th>Probability correct</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>&lt; 67%</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>&gt; 67%</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>-2</td>
<td>-6</td>
<td>&gt;80%</td>
</tr>
</tbody>
</table>

Figure 2: Gardner-Medwin’s optimal path to maximize score
In direct contrast to Gardner-Medwin, Davies [9] produced an MCQ scoring for 4 optional answers, where the students who demonstrate a high level of confidence in a correct choice should receive a greater reward, double that for high confidence in an incorrect answer, while penalising a student who demonstrated high levels of confidence for incorrect answers disproportionally less. However, like Gardner-Medwin, Davies also forgives the individual who declares low confidence in an incorrect answer by not penalising them at all.

2.3 Enhanced MCQ Assessment Incorporating Confidence Measurement (MCQCM)

Multiple Choice Questions with Confidence Measurement (MCQCM) is another online assessment tool developed and implemented by Farrell [7, 14] for computer science students, utilizing confidence measurement and subsequent penalties. The MCQCM was designed to address the areas of concerns previously identified, being to deter guessing, recognise partial knowledge and the place the control of learning into the hands of the student. The MCQCM was also designed to provide an entertaining learning experience, by incorporating an element of game play (waging, risk and gain) while providing the perception of fair play. The operational aspects of the MCQCM can be seen in Figure 3. The system requires the student to choose the correct answer(s) and identify the incorrect answers.

Figure 3: MCQCM interface, student selection.

A level of confidence is then registered for each option. Ref withheld considered it to be imperative that the MCQCM was easy to use, without placing too much cognitive demand on the user, ensuring that their efforts were placed on the question rather than the interface. It was also considered important that the scoring technique was simple, as opposed to Pollard’s [17] scoring (Table 1), placing the user in direct control of the results of their actions. On completion of the test the student receives an interactive graphical display of their results for each question to equip them for varying their study plan to address their shortcomings of their knowledge of the area being assessed, as in Figure 3 and 4.

The scoring method was adopted after extensive consideration to the previously discussed scoring mechanisms of Pollard [17], Paul [10], Gardner-Medwin [8] and Davies [9]. In order to enable an improved reporting of the student’s knowledge, a scoring method that proportionally rewarded and penalised a student for a correct and incorrect answer was the preferred choice. The main arguments for this decision was to develop a fairer system for the student that was in their direct control, while offering comparable expected outcomes to other scoring regimes. Ref withheld considers the MCQCM to offer a self-assessment experience that is honest, informative, and directional, whilst still providing a sense of fairness. The score is calculated dependent directly on their registered level of confidence for each option, using both positive and negative values. This incremental balanced scoring method is briefly explained in Table 3 with some simple examples.

Table 3: Rules and example of a score for a given scenario

<table>
<thead>
<tr>
<th>Confidence registered for an option</th>
<th>Example of score calculation from a registered confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>A high level of confidence for a correct answer for an option yields a high positive score</td>
<td>A confidence level of 90% for a correct answer yields a positive score of 9/10.</td>
</tr>
<tr>
<td>A high level of confidence for an incorrect answer for an option yields a negative score of equal value.</td>
<td>A confidence level of 90% for an incorrect answer yields a negative score of -9/10.</td>
</tr>
<tr>
<td>A low level of confidence for a correct answer for an option yields a low positive score.</td>
<td>A confidence level of 20% for a correct answer yields a positive score of +2/10</td>
</tr>
<tr>
<td>A low level of confidence for an incorrect answer for an option yields a negative score of equal value.</td>
<td>A confidence level of 20% for an incorrect answer yields a negative score of -2/10</td>
</tr>
</tbody>
</table>

Figure 4: Results summary to student’s answer.
The individual scores for each option allocated out of 10 are tallied to give a score for each of four questions, resulting in a score out of 40. Each score is displayed as a value from 0 to 10, or the negative equivalent. As an example, Table 4 demonstrates the resulting scores for a student’s answer to a question with a single correct answer as highlighted.

Table 4:  
Resulting score for registered level of confidence.

<table>
<thead>
<tr>
<th>Instructor’s Choice</th>
<th>A: i-</th>
<th>B: i++</th>
<th>C: i=1</th>
<th>D: i=i++</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student’s Choice</td>
<td>FALSE</td>
<td>TRUE</td>
<td>FALSE</td>
<td>FALSE</td>
</tr>
<tr>
<td>Correct or Incorrect</td>
<td>Incorrect</td>
<td>Correct</td>
<td>Correct</td>
<td>Correct</td>
</tr>
<tr>
<td>Confidence</td>
<td>65</td>
<td>90</td>
<td>100</td>
<td>92</td>
</tr>
<tr>
<td>Score</td>
<td>-6.5</td>
<td>9</td>
<td>10</td>
<td>9.2</td>
</tr>
</tbody>
</table>

The resulting final score for this question is calculated by the addition of the scores for each option:

$$E(X) = 0.25(1) + 0.75(0) = 0.25$$ (Supporting Guessing)

Ref withheld (2010) in their supporting the development and implementation of the MCQCM provides a series of $E(\text{Score})$ with appropriate graphs to demonstrate the possible outcomes. This simple demonstration shows that the incremental balanced scoring method is a proper scoring system, as Paul and Gardner-Medwin promote, as it rewards the participant proportionally to their level of stated knowledge, whilst permitting and encouraging the demonstration of partial knowledge.

3. Conclusion

This research introduces the MCQCM, an innovative assessment strategy for computer science students, designed to increase the level of feedback to the students. The MCQCM scoring was derived from reference to the previous extensive work of others with the objective to construct an appropriate scoring system that offers both fair and proportionally rewarding.

This research initially discussed Pollard’s work, an early pioneer of implementing an assessment strategy incorporating penalties, provided a purpose-designed solution for the Australian Mathematical Association National Mathematics Quiz, required to deter students from guessing. While it achieved its primary purpose of deterring guessing it has seen limited success in servicing general education as it relies on a complex scoring regime that is considered to be beyond the comprehension of the average tertiary student. In addition, being designed to identify the highest achieving students, which it does so eloquently, it seems to be a heavy handed approach for general purpose.

Following that, Paul’s [10] CBAA interactive assessment tool based on the triangle with each apex representing the option offered a truly engaging experience and has reported excellent results, with extensive use over the years. While the value of the CBAA is well recognized there are some who have reservations of the scoring system, considering it to be a complex logarithmic based solution, increasing the cognitive load on the student, possibly resulting in a negative effect on the final outcome and again moving away from placing the control in the student’s hands. Hence, the standard Multiple Choice Question (MCQ) with four options, one correct answer and no penalties for incorrect answers has the $E(\text{Score})$ calculated by

$$E(X) = 0.25(1) + 0.75(0) = 0.25$$ (Supporting Guessing)

Finally, the authors feel that the requirement of the student possessing a level of dexterity to adequately negotiate the regions with the cursor could disadvantage some students. A comparative analysis of the expected values of the incremental balanced scoring method to other more complicated methods demonstrate that the expected outcomes are not significantly different to warrant the additional cognitive load [14].
Davies’ [9] MCQ has achieved good outcomes and has been used extensively. Davies considers it to be of a supportive nature, as the minimal penalties for high levels of confidence for incorrect answers can assist struggling students, being a gentler reminder to the student that they are delusional in their perception of their knowledge. The authors feel that it just does not go far enough to deter guessing or in highlighting the seriousness of their lack of understanding.

Gardner-Medwin legitimately argued, supported by mathematical modeling and probability theory, a harsh penalizing high confidence (C=3) for an incorrect answer of -6. Some academics deem this too severe. Gardner-Medwin stands firm on his position stating that “We fail if we mark a lucky guess as if it were knowledge….We fail if we mark confident errors as if they were no worse than acknowledged ignorance”, then by saying “there is a reticence to apply penalties, which in his opinion is an irresponsible approach in educating students”. Academics have also registered concerns that this type of testing leads the student to develop strategies to do well, by minimizing their losses and maximizing gains, which Gardner-Medwin considers part of the learning process, in developing a strategic approach. The authors acknowledge the legitimacy of this scoring method but have concerns regarding the perception of fairness. Like the criticism of Pollard’s system, they feel that it has been designed for the higher achievers and might not transfer well to the more general body in education. Some academics question if their students are robust enough to deal with this level of honesty regarding their knowledge, or lack thereof.

Leveraging off this previous work Farrell [14] deemed it important that the MCQCM scoring calculations remain linear and simple to permit direct control by the student. Any distraction from the primary goal of correctly answering the question could have severe consequences for the student. Equally important was the need for the activity to be playful and engaging (laying a wager as a means of supporting your choice), fair and finally properly rewarding and motivational, in that scores would be rewarded for levels of confidence in correct answers.

Even though the incremental balanced negative scoring is based on a proper scoring system that addresses the area of guessing and recognises partial knowledge, it has been criticised for not being truly motivating. This criticism is valid in that the equivalent to Gardner-Medwin’s optimal path analysis encourages the option of refraining from answering when low confidence, yielding no penalty. However, the students also appreciate the fact that no gain can be made by abstaining from the activity, which is apparent from the implementation phase where it was reported that students infrequently chose to refrain from committing to answers.

References
[16] G. Krätzig and K. Arbuthnott, ”Metacognitive Learning: The Effect of Item-Specific Experience and Age
on Metamemory Calibration and Planning, Metacognition and Learning," vol. 4 2009, 125-144.