Constructing feedback for computer science MCQ wrong answers using semantic profiling (Discussion Paper)

JANE WAITE, Raspberry Pi Foundation, United Kingdom and University of Cambridge, United Kingdom
EIRINI KOLAITI, Raspberry Pi Foundation, United Kingdom
MEURIG THOMAS, University of Cambridge, United Kingdom
KARL MATON, University of Sydney, Australia

Computer science (CS) for K-12 students is a hard subject with abstract concepts to learn. Multiple-choice questions (MCQ) are often used to supplement classroom learning, with feedback from wrong answers playing a role in addressing misunderstandings. Semantics, a dimension of Legitimation Code Theory, a framework for understanding knowledge practices, has been suggested as a useful theory for reviewing and structuring CS learning events. In this discussion paper, we explore the use of 'semantic profiling' to improve feedback to wrong answers in MCQ for post-16 students studying SQL and relational databases. We describe the reflexive review process we developed and present the semantic profiles of two case studies of new feedback for wrong answers. New answer feedback to five questions was trialed in a pilot study with five students, and students self-reported the new feedback as useful. For example, students liked the metacognitive aspect of feedback that explained why answers were wrong or right and liked the generalised summaries. From our reflexive experience, we suggest reviewing MCQ questions and using semantic profiling for feedback can help MCQ authors and students develop their feedback literacy, particularly for creating learner 'feedforward' (take-away) opportunities. The approach we used has promise that we will build on, and we invite other researchers to explore and evaluate this approach.

CCS Concepts: • **Social and professional topics → K-12 education**.

Additional Key Words and Phrases: computing education, K-12 education, feedback literacy, SQL, databases

ACM Reference Format:
Jane Waite, Eirini Kolaiti, Meurig Thomas, and Karl Maton. 2023. Constructing feedback for computer science MCQ wrong answers using semantic profiling (Discussion Paper). In *23rd Koli Calling International Conference on Computing Education Research (Koli Calling '23), November 13–18, 2023, Koli, Finland*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3631802.3631803

1 INTRODUCTION

The teaching of computer science (CS) is challenging, involving abstract concepts that can be difficult to learn. For students aged 17 to 18 taking higher-level CS examinations, there is pressure to achieve good grades as they approach competitive applications for university. To achieve this, some teachers turn to online systems to supplement in-class teaching with multiple-choice questions (MCQ) that provide practice as well as formative and summative assessment of knowledge. Commentary from teachers and students using a student-facing CS education platform over the last four years, the period of its existence, has been that the MCQ feedback needs to be improved for wrong answers. This platform is an online learning tool that provides school-aged CS content and questions, including MCQs [31]. To address the request to improve feedback, especially as feedback has been suggested to be a powerful influencer on achievement [13], this pilot study has been conducted to investigate using concepts from the Semantics dimension of Legitimation Code Theory (LCT) [16] to improve online feedback to MCQ. LCT is useful for learning complex CS concepts in schools

Manuscript submitted to ACM

[5, 32] and has been used to analyse written feedback by tutors for undergraduate essay writing in English Studies [30], but as yet has not been explored for constructing MCQ feedback for wrong answers in CS.

2 RELATED WORK

Students studying at age 17-18 years old for formal CS examinations are often required to learn about databases and associated programming languages [8, 11, 14]. Specifically, in some countries, concepts such as normalisation, primary keys, foreign keys and referential integrity are learned about and how these are implemented in Structured Query Language (SQL) (e.g. [1]). Research related to the teaching of databases and SQL has focused on university settings, where this topic is a mainstay of undergraduate courses [29] with research including teaching approaches, student errors, concepts taught, easing teacher workload [29] and using automated marking tools [23]. For K-12 education, there is little research on teaching these topics, except the work by Grillenberger who has developed a simplified programming language for high-school students to develop database applications (EledSQL) [10] and provided a data literacy competency model including processes such as modeling and analyzing data [12]. Also, there has been work done with a small number of high-school and university students on types of SQL misconceptions [20], which will be important to address, as with when learning programming (e.g. [6]), including through feedback to students.

In a synthesis of over 800 meta-analyses of classroom teaching and learning research, Hattie reported feedback as one of the most "powerful influences on achievement" [13, p.173]. However, this referred to when teachers found out about students learning, e.g. what errors they had made. Also, Hattie reported that more often than not, feedback in classrooms, for example, from peers, was incorrect [13, p.4]. Feedback provided by technology might be expected to be more often correct than from peers. However, how useful it is to learners is likely related to the pedagogy being used in the type of learning activity [28, p.20, p.23]. Work on student 'feedback literacy' has described 'feedback literate' students as understanding and appreciating: the role of feedback in improving work, the active role needed by learners, and the different processes they might use to effectively action feedback [4]. Feedback has been categorized into four types, each with different student processes to use: a) Telling, a uni-directional transmission of 'correct' answers assuming a passive role for the student; b) Guiding, in which the students are being pointed in the right direction so that they may learn by applying knowledge to practice; c) Developing understanding, creating meaningful abstractions which require students to be active in their construction or adjustment of knowledge structures; d) Deliberately opening up a different perspective, which requires students to be actively engaged in interpreting and evaluating knowledge [19].

Technology can support how concepts are explained, how assessment is delivered, and provide timely learner feedback [28], including using online multiple-choice questions (MCQ), which have become popular. Guidance on MCQ construction includes a) Keeping the question structure simple, and ensuring the question has a clear, suitable learning objective [3]; b) Only common errors, misconceptions and plausible answers should be used as distractors not fabricated, false answers that are out of context [3]; c) Limiting answers to plausible options, typically three of which one is correct [26]; d) If seeking to increase the difficulty of MCQs, it is important to scaffold learning and provide feedback that allows students to correct their errors [3]; e) Students can learn an incorrect answer by selecting it; therefore the wrong answer should be challenged [3]. However, guidance on MCQ feedback construction is less clear; research on general e-assessment feedback has found high variability of impact on student outcomes and limited research [7].

Legitimation Code Theory (LCT) offers a conceptual toolkit that can be used to interpret the various dimensions of educational practice [16]. Semantics is one dimension of LCT and involves the concepts of semantic gravity (SG) and semantic density (SD). Semantic gravity expresses how closely meaning relates to its context and can be stronger or weaker along a continuum of strengths; for example, semantic gravity would be stronger if an explanation was

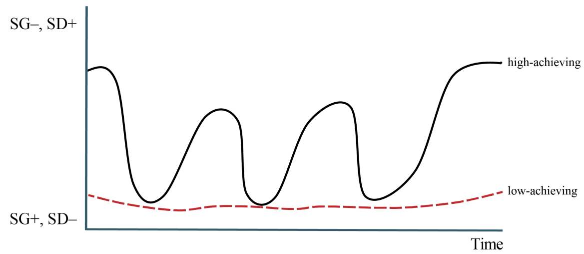

Fig. 1. Example semantic profiles of two students' essays [16, p.19]

dependent on a specific context and weaker, say if there was no context in a generalised definition. Semantic density refers to the complexity of meaning and can also be stronger/weaker. Semantic density is stronger when, for example, a claim condensed many meanings within it and weaker if that claim involved relatively fewer meanings. These strengths can be represented on a 'semantic profile' (Figure 1) where the y-axis shows strengths of semantic gravity and semantic density and the x-axis represents time or text-time [16]. Some profiles look like a flat-line, where the strengths stay the same over time, while others may depict 'semantic waves', where strengths move up and down. For example, Figure 1 shows the profiles of two students' essays: the low-achieving student's profile is a low flat-line, whereas the high-achieving student's essay comprises waves or recurrent moves down and up [16, p19]. Moves down are when complex meanings are simplified (weakening semantic density) and grounding ideas in concrete examples (strengthening semantic gravity). Moves up are when more complex meanings are built and meanings generalised and abstracted and concrete examples are linked back to abstract concepts and terms [16]. The idea of balancing abstract and concrete feedback is not new in CS (e.g. [15]), but semantic profiling brings a formalisation to this. Semantic profiling has been applied to unplugged CS activities and introductory programming tasks with the suggestion that semantic waves can enrich learning experiences and improve knowledge acquisition [5, 32]. Focusing on assessment, these LCT concepts have been used to analyse undergraduate chemistry question construction [27] and to review undergraduate students' free-text answers to physics questions [9]. In the teaching of undergraduate writing, semantic gravity has been used to investigate tutors' 'feedback literacy', with the finding that written feedback was often too context-bound and not generalised so did not reach up the semantic profile, hence students could not easily use it to 'feed-forward' into their next writing activity that would be in a new context [30]. It is this move upwards, conveying ideas that students can take out of the specific context and into new contexts, that we focus on here.

This study uses semantic profiling as a framework for developing structured online MCQ feedback presented to students after they have selected a right or wrong answer on a specific learning platform. The platform has been designed to be used in classrooms, by teachers to supplement traditional CS teaching [31]. The platform presents various question types all with a practice pedagogy of multiple tries with tailored feedback. When attempting questions, if the student selects a correct or an incorrect answer, feedback is provided instead of the correct answer. Students can attempt a question as many times as they like. Being able to practice problem-solving with immediate feedback scaffolds learning, provides low-stakes formative self-assessment, promoting self-efficacy, and self-regulation [21, 31].

3 DESIGN OF THE STUDY

In this pilot study, we focused on the feedback provided to MCQ answers investigating how semantic profiling might be used to review and construct MCQ answer feedback, and what students might think of such feedback. There were four phases of the study. Ethics approval was obtained for working with participants and handling their data. The

Vanessa is an artist that creates paintings. Each original painting is displayed in her gallery. She doesn't sell the original work but customers can purchase posters of the paintings. There is a limit to how many posters are produced for each painting. For example, for the 'Constellations' painting there are only 60 posters available. To keep track of how many posters have been sold for each painting, each poster is numbered.

Painting Name	Poster Limit
Nature disturbed	75
Constellations	60

Vanessa uses a relational database to store information about the sales. This includes details about the paintings and the posters she has produced, her customers, and the sales of the posters. The information is modelled using three entities: `Painting`, `Customer`, and `Sale`. The description in standard notation for the entities is as follows:

```
Painting(PaintingId, Name, Description, ProductionYear,
PosterLimit)
Customer(CustomerId, LastName, FirstName, Email)
Sale(PaintingId, PosterNumber, CustomerId, SaleDate, Cost)
```

What is the relationship between the entities `Painting` and `Sale`?

- C - Many-to-many
- B - One-to-many
- A - One-to-one

Fig. 2. Question 1

Suhreena wants to create a program for her A level project. The program will allow her to run polls that the students can participate in using the school portal.

- Each poll consists of one question and any number of response options. Each response option belongs to a specific poll.
- Each student can only vote for one option in each poll. Once a vote is submitted, it can't be changed.
- When a poll is over, it can be set as inactive.

Suhreena wants to use a relational database to store information about the polls. The information is modelled using four entities `Student`, `Poll`, `Option` and `Vote`. Each entity is implemented in the database using a table. The description in standard notation for the entities is as follows:

```
Student(StudentId, FirstName, LastName, YearGroup)
Poll(PollId, QuestionText, IsActive, PublishedDate)
Option(OptionId, PollId, OptionText)
Vote(StudentId, OptionId, SubmittedDate, SubmittedTime)
```

Suhreena wants to produce a list with information on all the polls that are active in the database. For each poll, she wants the poll question and options, and the date it was published. Which one of the below `SELECT` statements is correct?

B -

```
SELECT Poll.QuestionText, Option.OptionText, Poll.PublishedDate
FROM Poll, Option
WHERE Poll.PollId = Option.PollId
AND Poll.IsActive = TRUE
```

A -

```
SELECT Poll.PollId, Poll.QuestionText, Poll.PublishedDate
FROM Poll
WHERE Poll.IsActive = TRUE
```

C -

```
SELECT Poll.QuestionText, Option.OptionText, Poll.PublishedDate
FROM Poll, Option
WHERE Poll.PollId = Option.OptionId
AND Poll.IsActive = TRUE
```

Fig. 3. Question 2

research question for this study was: *RQ In what ways was MCQ feedback changed using semantic profiling, with the aim of improving it, and how did students respond to the new feedback?*

Phase 1: Question selection Rather than creating new questions, existing questions were taken, reviewed and changed. Questions were selected from the platform. To reduce any cognitive load that might result from participants as they swapped topics, two aligned topics were chosen to select questions of databases and SQL. These topics were chosen as they were commonly studied topics for the age group yet had limited research conducted (Section 2), therefore leading to a longer-term contribution. Three questions were picked from the existing question set of eleven for databases and two from sixteen for SQL. Question selection was relatively arbitrary, apart from giving a breadth of concepts.

Phase 2: Developing new versions of the answer feedback The second author produced an initial version of the updated feedback, and then the first author, who is experienced in semantic profiling, profiled the initial version. Together they reviewed the profile and updated the feedback to complete the new version. The objective was to produce movements up the profile, from very context-bound and simpler meanings to more context-independent and complex

Reflexive prompts	Question 1	Question 2
1. About the learning objectives		
1.1 What is the learning objective?	Work out the cardinality of the relationship between two entities given a scenario and data requirements.	Construct a SQL query selecting data from two tables using an appropriate join and filters.
1.2 Categorise by revised Bloom's level.	apply and analyse	apply and analyse
1.3 Categorise by type of knowledge needed.	declarative, procedural, conditional	declarative, procedural, conditional
2.0 About concepts and contexts		
2.1 Identify core technical language e.g. core concepts and abstract/high-level terms	e.g. entity, many-to-many, one-to-one, one-to-many, instance	SQL terms (e.g. SELECT, FROM, WHERE), primary key, foreign key, join
2.2 Remove any superfluous terms	e.g. "high quality signed copies"	No change needed
2.3 Remove non-related distractors	One wrong answer removed	All answers were functional
2.4 Separate vocabulary of dependent concepts	ER modelling terms separated from database implementation terms	ER modelling terms separated from database implementation terms
2.4 Make culturally relevant	e.g. "collage" changed to "painting"	No change needed
3.0 About feedback knowledge representations		
3.1 Add a different representation	ER diagram	Database tables with example data
3.2 Highlight key terms.	bold, bullet points, different fonts, coloured text, signalled important aspect with starter phrase "Notice that..."	all-caps, camel case, different fonts, coloured text, signalled important aspect with starter phrase "Notice that..."
4.0 Construct the feedback for each question answer considering the semantic profile.		

Table 1. Reflexive prompts and commentary for each case study question and feedback

meanings, ending with a relatively decontextualised summary of the concepts. Throughout this process, a reflexive journal was kept to log the discussions and decisions made by the researchers.

Phase 3: Student workshop and qualitative data analysis Having created the feedback for the five selected questions, these were shared with five students aged between 17 and 18 years old in an online workshop. The workshop was not associated with a school activity; rather, it was a separate research context and run by the research team. Recruitment for the pilot was purposive; students were invited by email from an existing voluntary student panel (related to the online platform) to take part in the research workshop. Students are not paid to be on the student panel. The online workshop took one hour and involved the students trialling the five questions with the new answer feedback in the online platform. The students were told that new questions were being explored for the online platform and were asked to look at the feedback for each answer (right or wrong) for each of the five questions. They were asked to complete an online survey as they went along, noting what they thought of the feedback. Students were asked "What is your view of the changed feedback for wrong answers?", "What is your view of having feedback for correct answers?" and "How can we help your teachers to use the new functionality effectively?"

Data from the student surveys were analysed using an inductive/deductive qualitative method [18] drawing from reflexive thematic analysis [2]. Reflexive thematic analysis requires the researchers to spend time actively reading and re-reading the data, finding repeated patterns of meaning within the text. Having uploaded the survey responses to NVivo, the second author reviewed the data looking for themes, and then discussed these with the first author agreeing on initial codes. The second author then coded up the data from initial ideas, inductively and deductively splitting and merging themes. The coding was then reviewed with the first author, and consensus was achieved for all coding. To increase the trustworthiness of the method, the first and second authors journaled their activities, including thought processes, and decisions and continually collaborated at each phase [24].

Phase 4: Case studies Two sample profiles of updated feedback for wrong answers were selected and reviewed with the fourth author, an expert in the field of semantic profiling. The objective of this was to verify the profiling process, discuss the resultant profiles as part of the reflexive approach and develop the related discussion. The selected profiles became case studies and are called Question 1 and Question 2. As well as sharing the detail of the profile of these wrong answers, we have described the commentary of our reflexive prompts for the questions (see Table 1).

As we had a small number of questions ($n = 5$) from a limited concept base (databases and SQL) and the sample size of participants was small ($n = 5$) and were purposively recruited, generalisation should be treated with caution.

4 FINDINGS

As we reviewed and adapted the answer feedback to control the gradual contextualization and simplifying of ideas (moving down the profile) or decontextualisation and complexifying of ideas (moving up the profile) to create a smooth but deep relative difference between the bottom and top of the profile for each feedback, we developed a set of reflexive prompts to help us think about the learning objectives, concepts and contexts and knowledge representations (Table 1).

4.1 Case studies

We selected two questions as case studies. One case study question (Question 1), was an entity-relationship question (Figure 2) and the other was a SQL question (Question 2) (Figure 3).

The original feedback for the wrong answer C for Question 1 was *"A sale is for a copy of a collage, but a collage will have many sold copies"*. This sits toward the bottom of the profile (see Figure 1) as it uses everyday language in a specific context. For Question 2, answer option C, the original feedback was: *"This SQL SELECT statement will produce a list of all the active polls but will not include the options of each poll"*. This sits a little further up the profiles as does involve more complex language alongside simpler everyday language but it remains situated in a specific context.

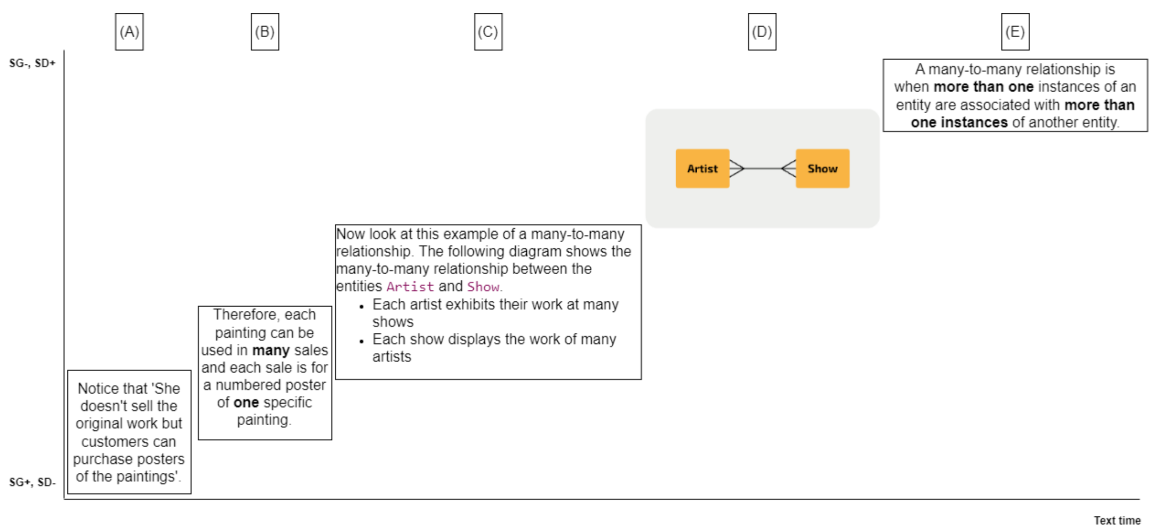

Fig. 4. Semantic profile for the feedback for wrong answer C for Question 1

The commentary of our responses to the reflexive prompts for the two case study questions is shown in Table 1. The new feedback and its semantic profile for option C, a wrong answer, for Question 1 are shown in (Figure 4). This feedback profile for Question 1 moves upwards or is what LCT terms an 'up escalator' profile. We have broken the feedback into five parts (A-E) shown verbatim in Figure 4. Firstly, (A), the reader's attention is drawn to a quote from the question that is important to address the wrong answer chosen, but using everyday language and situated in the question context. Then, (B), the quote is summarised and interpreted and primes the readers about key technical vocabulary using bold to draw attention - this shifts up a little through moving away from the question context and introducing complex terms. (C) shifts up again by using another context in the scenario, exemplifying a many-to-many relationship, and bringing in more technical terms and concepts (foregrounded through the use of bullet points, colours, and fonts). (D) shifts upwards: most of the context is stripped away in favour of an abstract Entity-Relationship diagram that condenses many meanings such as cardinality. Finally, (E) is a single 'take-away' message in which all the context from the question has been removed and meanings are technical terms setting out a standard technical but complex definition of a many-to-many relationship.

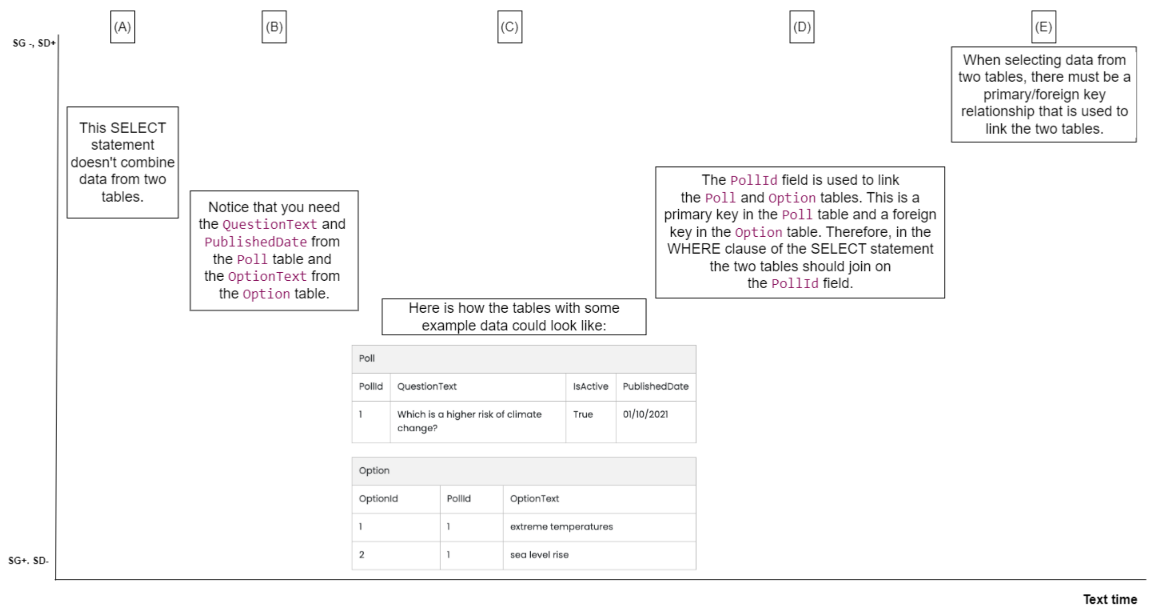

Fig. 5. Semantic profile for the feedback for wrong answer C for Question 2

The new feedback and its semantic profile for option C for Question 2 are shown in Figure 5. The profile this feedback reveals is a semantic wave that starts and ends high. We have broken the feedback into five parts (A–E), shown verbatim in Figure 5. The feedback begins (A) by referring to the incorrect answer by using the technical term "SELECT statement" and without drawing on the framing context of the question. It is thus relatively decontextualized (weaker semantic gravity) and complex (stronger semantic density). In (B), the attention of the reader is drawn to database tables and fields, using technical vocabulary – QuestionText, PublishedDate, Poll, OptionText and Option – taken directly from the question, including using the same font and colour as they appear in the question. Thus, semantic gravity is strengthened as the feedback leans into the specifics of the question, while semantic density remains relatively strong. In (C), a table is introduced with data to exemplify the feedback. This is mostly lacking highly technical terms, uses relatively simpler language, does not describe complex processes or relationships and is grounded by examples: stronger semantic gravity and weaker semantic density. In (D), we return to a similar kind of feedback to (B) in terms of the semantic profile. Here, the feedback again directly uses terms (including font and colour), such as technical SQL commands and table and field names, from the question, as well as further concepts ('primary key', 'foreign key', and 'WHERE clause'). The feedback also outlines relations among these terms (e.g. which field is used to join the database tables). Thus, it is moving up from the simpler example of the table in (C), increasing the abstractness and complexity of the knowledge being expressed. The final feedback step (E) comprises a generalised technical statement shorn of any mention of the context of the question. This 'take-away' statement is a generalisation that condenses all the previous meanings for all actions of this kind: weaker semantic gravity and stronger semantic density.

These case study question feedback options were selected as we had created different profiles: an up escalator, and a semantic wave profile. In either case, we could have constructed any type of profile, a flat-line, down, or up escalator, single wave, or multiple waves. We chose to create feedback text with different profiles, each of which moved towards ending high on the profile with a contextualised and technical 'take-away' message which has been suggested to improve feedback [30] and learning experiences [16].

4.2 Student views on changes

Students' free-text workshop survey answers, amounting to over 1,400 words, were analysed to develop themes (Table 2).

Theme (all expressed as positive views)	Student ids
1 Explains why each option is right or wrong.	S1, S2, S3, S4, S5
2 Provides detailed, thorough, deeper explanation.	S1, S4
3 Potential misconceptions are addressed.	S3, S4
4 Highlights what is important.	S1, S4
5 Explains the approach to solve the question.	S4, S5
6 Signposts key knowledge in various representations.	S1, S4
7 Provides a generalised summary.	S3, S4

Table 2. Themes of student responses to new feedback

Theme 1: Students found that clarifying why an option is right or wrong was very helpful. *"I find this feature useful because instead of saying the answer is correct, it also explains why the answer is correct which can be especially useful if the student guessed the answer."* [Student S1] *"I think it is helpful because it provides very detailed feedback as it gives an explanation of the terms mentioned in the question and then it uses that to explain why your answer is wrong."* [Student S4]

Theme 2: Two students commented that the increased level of detail in the new explanations was helpful: *"The feedback for answer B was helpful because it simplified why the two keys were foreign keys and from what table they are primary keys in."* [Student S4]

Theme 3: Two students viewed the new feedback as helpful to address potential misconceptions. *"I like how the information is related to the wrong answer rather than the right answer, (e.g. about composite keys rather than foreign keys) because there must have been some misunderstanding about composite keys in this case, but talking too much about foreign keys is just giving away the right answer."* [Student S3]

Theme 4: With regard to highlighting what information was important, two students mentioned this, e.g. *"For Question 1, the feedback for answer C was helpful because it boldened the important information that could have caused misconceptions."* [Student S4]

Theme 5: Two students said feedback was helpful as it gave guidance on how to approach solving the question: *"For Question 5, the feedback for answer C was helpful because it showed what the result of the wrong SELECT statement would be and it clearly showed the reasons on how the student should have approached the problem."* [Student S4]

Theme 6: Two students noticed different representations of knowledge, e.g. *"the feedback is useful because of the use of tables and using bold to highlight key terms to help students get the correct answer."* [Student S1]

Theme 7: Two students explained how they liked the generalised summary of a concept: *"It is really nice to have a concise statement about the answer to the question. I like that this statement is different to the answer of the question - it is just summarising foreign keys in general, for instance, rather than giving the answer again."* [Student S3] *"For question 2, the feedback for answer C was really helpful as I struggle with relationships between entities. I like how it said why the answer was wrong, gave an example of a many-to-many relationship and how it reinstated what a many-to-many relationship was generally."* [Student S4]

5 DISCUSSION

Our research question for this study was: *In what ways was MCQ feedback changed using semantic profiling, with the aim of improving it, and how did students respond to the new feedback?*

We developed a set of reflexive prompts that helped us prepare for profiling the feedback for each question. This gave us a consistent approach to the process as we built up an understanding of how to work with the concepts from LCT to provide a relative change in the semantic gravity and semantic density of ideas and finish on a generalised statement of the relevant concept. Using our reflexive prompts, we first checked that questions and answers met suggested guidelines for multiple choice questions, e.g., that all distractors were plausible etc. (Section 2). In general, the questions already met the requirements for well-structured MCQ, few changes were needed. Rather the changes made were in constructing feedback in the structure of a semantic wave or up escalator for each answer.

One significant aspect of the feedback was the addition of a detailed explanation of why the relevant answer was either right or wrong. Students liked this aspect of the new feedback and that related misconceptions were directly addressed. In research on MCQs, it has been recognised that misconceptions can be re-inforced by students selecting wrong answers (despite the answer then being marked as wrong) and that feedback is needed to tackle the negative impact [3]. This issue was directly addressed in the new feedback that we created, as the reasons why an option was right or wrong was explained. This change provided meta-knowledge about answers, giving students a clear signal that they can use this knowledge, building students' 'feedback literacy' as they take an active role in working with feedback to adjust their knowledge [19] through the explicit and detailed explanation about the answers.

As part of the profiling, we focused on core technical terminology and how to carefully control its introduction and the complexity of meaning of these terms. In our review of the two case study profiles, we noticed that we primed the students to prepare them about forthcoming technical detail. The priming was often through typographical cues, making them bold, or in italics, in a different colour, or even through bullet points. We also started to use phrases such as "Notice that..." to draw attention to significant elements of the question. Then with each step of the structured explanation, the context of the question was stripped away as technical vocabulary was gradually introduced. For example, the technical term 'many-to-many' was primed by using bold typography of the word 'many' in the context of the question. Then a 'many-to-many relationship' was explained using a different example (within the same question scenario), and the word 'many' was signalled as being important in bullet points to exemplify the relationship between two entities. Students remarked they liked having what was important signalled and they noticed the different formats used to do this. Research on typography cues and multimedia in reading comprehension indicates cues focus the reader's attention [17]. However, it is less clear whether readers remember or understand terms more when being cued, with indication there may be learner illusions with increased confidence to understand, but this does not always lead to long-term knowledge building [22]. Further research is needed to explore strategies such as priming or signaling.

Different formats of knowledge representation were also used more explicitly. For the second case study (Question 2 case), we exemplified a potential misconception that may have contributed to the wrong answer by directly connecting to the scenario of the question using a table of example data. This presentation of the data can be categorised as relatively strong semantic gravity (context-dependence). Students remarked they found such exemplification useful. For Question 1, we added an Entity-Relationship (ER) diagram. This knowledge representation acts very differently to the table of data. To interpret ER diagrams, a person has to understand the meaning of the shapes and lines, for example, the splayed lines (i.e. crow's foot) depict a cardinality of 'many'. There is much condensation of knowledge in this abstract representation (weaker semantic gravity and stronger semantic density). Again, students noted the use of diagrams as useful, but here compared to the table of data, the knowledge representation is very different with respect to the relationship to semantic gravity and semantic density. In research on what media affects learning and how (e.g. [17]), using semantic gravity and semantic density could help to analyse in more detail the differences between media.

Generally, in reviewing the semantic profiles, we noted two 'frames' for each question. Firstly, the context in which the question was situated, such as recording sales of posters (Question 1), and secondly, the type of learning activity - using MCQs. The MCQ learning activity influences the type of feedback [19] given and will be different compared to other ways of learning (such as peer feedback during pair programming). In MCQ learning activities, there may be more telling and some guiding, but perhaps less opportunity to develop understanding and open up new views. Similarly, feedback literacy [4] for MCQs will likely differ from other learning activity types. Using MCQs, students may play the percentages to figure out how to answer a question by spotting patterns. If a question stem is changed or the question options are reordered, students may find earlier MCQ feedback hard to apply effectively. We argue that by creating detailed, rich, mini-learning feedback experiences, students' views of MCQ for learning may change, raising the value of MCQ for formative assessment. As part of this, semantic profiling provides MCQ authors, teachers and students with ingredients to unpick knowledge building in feedback and to consider the role of students in using the components of feedback. For example, a key objective of our new feedback was to leave students with a technical and generalisable 'take-away' message high up on the profile that could to be applied to other MCQ questions and even used outside the setting of MCQs. This approach of ending with a generalised view was noticed and appreciated by students. Further research is needed to discover if such 'take-away' endings effectively support students' 'feedback literacy' [4] to adjust their understanding and to 'feed-forward' [30] to use in new other learning activities.

6 CONCLUSION AND NEXT STEPS

In this pilot study, we explored the use of semantic profiling to improve feedback to wrong answers in multiple-choice questions to post-16 students studying SQL and relational databases. We presented the reflexive prompts that we developed for feedback creation and the semantic profiles of two case studies of new wrong answers. New answer feedback was piloted with a small group of students, and the changes were welcomed. Students particularly liked how the feedback explained why answers were right or wrong. In the discussion, we drew upon research on feedback literacy to reflect on how semantic gravity and semantic density provide the ingredients to improve feedback literacy for students and content authors. Using the platform Ada Computer Science [25], we plan to trial the changed feedback through A/B testing, to compare different profiles and conduct further qualitative work with a wider classroom population. We invite other researchers to explore these opportunities in different contexts. As a final point, machine learning is starting to be used to create tutoring applications that provide student feedback. Whether large language models can be trained using data with specific semantic profiles remains to be seen, but certainly, investigating feedback literacy with semantic profiling for students, teachers and content writers seems an important avenue for further study.

REFERENCES

[1] AQA Examination Board. 2019. AS and A-level Computer Science Specification: 4:10 Fundamentals of databases. https://www.aqa.org.uk/subjects/computer-science-and-it/as-and-a-level/computer-science-7516-7517/subject-content-a-level/fundamentals-of-databases

[2] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport and health* 11, 4 (2019), 589–597.

[3] Andrew C. Butler. 2018. Multiple-Choice Testing in Education: Are the Best Practices for Assessment Also Good for Learning? *Journal of Applied Research in Memory and Cognition* 7, 3 (2018), 323–331. https://doi.org/10.1016/j.jarmac.2018.07.002

[4] David Carless and David Boud. 2018. The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education* 43, 8 (2018), 1315–1325. https://doi.org/10.1080/02602938.2018.1463354 arXiv:https://doi.org/10.1080/02602938.2018.1463354

[5] Paul Curzon, Jane Waite, Karl Maton, and James Donohue. 2020. Using semantic waves to analyse the effectiveness of unplugged computing activities. In *Proceedings of the 15th Workshop on Primary and Secondary Computing Education*. ACM, Virtual Event Germany, 1–10. https://doi.org/10.1145/3421590.3421606

[6] Rodrigo Duran, Juha Sorva, and Otto Seppälä. 2021. Rules of Program Behavior. *ACM Trans. Comput. Educ.* 21, 4, Article 33 (nov 2021), 37 pages. https://doi.org/10.1145/3469128

[7] Carol Evans. 2013. Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research* 83, 1 (March 2013), 70–120.

[8] Katrina Falkner, Sue Sentance, Rebecca Vivian, Sarah Barksdale, Leonard Busuttil, Elizabeth Cole, Christine Liebe, Francesco Maiorana, Monica M. McGill, and Keith Quille. 2019. An International Study Piloting the MEasuring TeacheR Enacted Computing Curriculum (METRECC) Instrument. In *Proceedings of the Working Group Reports on Innovation and Technology in Computer Science Education* (Aberdeen, Scotland Uk) *(ITiCSE-WGR '19)*. Association for Computing Machinery, New York, NY, USA, 111–142. https://doi.org/10.1145/3344429.3372505

[9] Helen Georgiou, Karl Maton, and Manjula Sharma. 2014. Recovering Knowledge for Science Education Research: Exploring the "Icarus Effect" in Student Work. *Canadian Journal of Science, Mathematics and Technology Education* 14, 3 (2014), 252–268.

[10] Andreas Grillenberger and Torsten Brinda. 2012. EledSQL: A New Web-Based Learning Environment for Teaching Databases and SQL at Secondary School Level. In *Proceedings of the 7th Workshop in Primary and Secondary Computing Education* (Hamburg, Germany) *(WiPSCE '12)*. Association for Computing Machinery, New York, NY, USA, 101–104. https://doi.org/10.1145/2481449.2481474

[11] Andreas Grillenberger and Ralf Romeike. 2014. Big Data – Challenges for Computer Science Education. In *Informatics in Schools. Teaching and Learning Perspectives*, Yasemin Gülbahar and Erinç Karataş (Eds.). Springer International Publishing, Cham, 29–40.

[12] Andreas Grillenberger and Ralf Romeike. 2018. Developing a Theoretically Founded Data Literacy Competency Model. In *Proceedings of the 13th Workshop in Primary and Secondary Computing Education* (Potsdam, Germany) *(WiPSCE '18)*. Association for Computing Machinery, New York, NY, USA, Article 9, 10 pages. https://doi.org/10.1145/3265757.3265766

[13] John Hattie. 2010. *Visible learning: a synthesis of over 800 meta-analyses relating to achievement* (reprinted ed.). Routledge, London.

[14] Peter Hubwieser, Michail N. Giannakos, Marc Berges, Torsten Brinda, Ira Diethelm, Johannes Magenheim, Yogendra Pal, Jana Jackova, and Egle Jasute. 2015. A Global Snapshot of Computer Science Education in K-12 Schools. In *Proceedings of the 2015 ITiCSE on Working Group Reports* (Vilnius, Lithuania) *(ITICSE-WGR '15)*. Association for Computing Machinery, New York, NY, USA, 65–83. https://doi.org/10.1145/2858796.2858799

[15] Petri Ihantola, Ville Karavirta, and Otto Seppälä. 2011. Automated Visual Feedback from Programming Assignments. (2011), 87–95. http://urn.fi/URN:NBN:fi:aalto-201609013232

[16] Karl Maton. 2013. Making semantic waves: A key to cumulative knowledge-building. *Linguistics and Education* 24, 1 (April 2013), 8–22.

[17] Richard E Mayer. 2012. *Multimedia Learning* (2 ed.). Cambridge University Press, Cambridge, England.

[18] Philipp Mayring. 2000. Qualitative Content Analysis. *Forum: Qualitative Social Research* Vol 1 (June 2000), No 2 (2000): Qualitative Methods in Various Disciplines I: Psychology. https://doi.org/10.17169/FQS-1.2.1089 Publisher: Forum: Qualitative Social Research.

[19] Angela J. McLean, Carol H. Bond, and Helen D. Nicholson. 2015. An anatomy of feedback: a phenomenographic investigation of undergraduate students' conceptions of feedback. *Studies in Higher Education* 40, 5 (2015), 921–932.

[20] Daphne Miedema, Efthimia Aivaloglou, and George Fletcher. 2021. Identifying SQL Misconceptions of Novices: Findings from a Think-Aloud Study. In *Proceedings of the 17th ACM Conference on International Computing Education Research* (Virtual Event, USA) *(ICER 2021)*. Association for Computing Machinery, New York, NY, USA, 355–367. https://doi.org/10.1145/3446871.3469759

[21] James Moore. 2020. Mark it yourself, it's good for you! A case study examining the impact of self-assessment on student learning and motivation in a year 12 physics class studying energy. *Journal of Trainee Teacher Education Research* 11 (2020), 38.

[22] Michael L. Mueller, John Dunlosky, Sarah K. Tauber, and Matthew G. Rhodes. 2014. The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people's beliefs about memory? *Journal of Memory and Language* 70 (2014), 1–12. https://doi.org/10.1016/j.jml.2013.09.007

[23] Sidhidatri Nayak, Reshu Agarwal, and Sunil Kumar Khatri. 2022. Review of Automated Assessment Tools for grading student SQL queries. In *2022 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, Coimbatore, India, 1–4. https://doi.org/10.1109/ICCCI54379.2022.9740799

[24] Lorelli S. Nowell, Jill M. Norris, Deborah E. White, and Nancy J. Moules. 2017. Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *International Journal of Qualitative Methods* 16, 1 (Dec. 2017), 1609406917733847. Publisher: SAGE Publications Inc.

[25] Raspberry Pi Foundation and the University of Cambridge. 2023. *Ada Computer Science*. https://adacomputerscience.org/

[26] Michael C. Rodriguez. 2005. Three Options Are Optimal for Multiple-Choice Items: A Meta-Analysis of 80 Years of Research. *Educational Measurement: Issues and Practice* 24, 2 (June 2005), 3–13. https://doi.org/10.1111/j.1745-3992.2005.00006.x

- [27] Ilse Rootman-le Grange and Margaret A. L. Blackie. 2018. Assessing assessment: in pursuit of meaningful learning. *Chemistry Education Research and Practice* 19, 2 (2018), 484–490. https://doi.org/10.1039/C7RP00191F
- [28] Eleanor Stringer, Cathy Lewin, and Robbie Coleman. 2019. *Using digital technology to improve learning*. Guidance Report. EEF, London, UK. https://educationendowmentfoundation.org.uk/public/files/Publications/digitalTech/EEF_Digital_Technology_Guidance_Report.pdf
- [29] Toni Taipalus and Ville Seppänen. 2020. SQL Education: A Systematic Mapping Study and Future Research Agenda. *ACM Trans. Comput. Educ.* 20, 3, Article 20 (aug 2020), 33 pages. https://doi.org/10.1145/3398377
- [30] Martina van Heerden. 2021. (How) do written comments feed-forward? A translation device for developing tutors' feedback-giving literacy. *Innovations in Education and Teaching International* 58, 5 (2021), 555–564. https://doi.org/10.1080/14703297.2020.1788411
- [31] Jane Waite, Andrea Franceschini, Sue Sentance, Mathew Patterson, and James Sharkey. 2021. An Online Platform for Teaching Upper Secondary School Computer Science. In *Proceedings of the 2021 Conference on United Kingdom & Ireland Computing Education Research* (Glasgow, United Kingdom) *(UKICER '21)*. Association for Computing Machinery, New York, NY, USA, Article 3, 7 pages. https://doi.org/10.1145/3481282.3481287
- [32] Jane Waite, Karl Maton, Paul Curzon, and Lucinda Tuttiett. 2019. Unplugged Computing and Semantic Waves: Analysing Crazy Characters. In *Proceedings of the 1st UK & Ireland Computing Education Research Conference on - UKICER*. ACM Press, Canterbury, United Kingdom, 1–7. https://doi.org/10.1145/3351287.3351291