Investigating the relation between students' reported self-confidence levels when answering questions and their learning behaviour

Alvaro Martin-Angulo

April 11, 2022

1 Overview

T he process of education is a very complex one and is one that is constantly evolving. One of the ways of analysing student learning behaviour in hopes to improve the educational system is by asking students to report their confidence level when answering multiple-choice questions.

Data from the EEdi educational platform was used to analyse in what ways the confidence levels reported when answering multiple-choice mathematics questions related to the learning behaviour of students. It was found through a logistic regression that the higher the confidence level reported for a question, the more likely it was to be answered correctly. Overconfident and under-confident responses were also analysed, and there were significantly more overconfident responses to questions (73771) than under-confident responses (16361).

Further analysis was done of the overconfident responses, showing that in some of the questions, a very large percentage of the total answers were overconfident responses (as high as 85.7%). This suggests that there are misconceptions in certain areas of students' learning as they are confident in their answers yet they are incorrect. Comparing the higher level of subjects examined in the questions, Numbers was the subject with the most misconceptions, followed by Geometry and Measure and lastly Algebra.

The age of the person at the time of answering a question was compared to the average confidence level reported to see if any correlation aroused between age and reported confidence level.

2 Introduction

Context and motivation This study aims to find information about student behaviour in hopes to obtain valuable results to improve the educational system. This area of study is particularly interesting due to its abstract nature - every student is different, and the psychology of each individual can have a big impact on reporting confidence levels when answering a question.

Furthermore, analysing the nature of questions and how well they were answered in relation to the confidence that was reported for that question may give information about misunderstandings in certain fields of the educational system.

Previous work A similar study was conducted on a database containing information of confidence on 20 multiple choice questions answered by 104 third-year dental students. The study concluded that more students were misinformed (22%) than uninformed (8%) [2]

Another study was conducted on 81 medical students on neurology multiple-choice exam [3]. In this case, in contrast to [2], the study concluded that there were more uninformed students than misinformed.

Surprisingly, these studies reached contrasting conclusions, perhaps due to the area of research of the questions. Investigating mathematics questions will be interesting as it is a completely different field of study.

Objectives When studying this data, several questions come to mind regarding the behaviour of students i.e., Even though a student may report that they might not be confident at all in their answer, will they be more likely to get a correct answer than if it were to be selected randomly? (This would imply that students make an educated guess despite not knowing anything about a question).

As a more general question, it would be interesting to find how much confidence levels relate to getting a correct answer. Other categorical variables such as age when answering a question or gender may also be interesting to analyse to see if they correlate with the level of confidence reported.

In terms of behaviour towards certain questions, levels of confidence for a question and how well the question was answered will be analysed to find questions with common misconceptions.

3 Data

Data provenance The data comes from the EEdi NeurIPS Education challenge 2020 [4]. A subset of the data has been provided by the University of Edinburgh as files to work on for the project. The data is licensed under a Creative Commons Attribution- NoCommercial- NoDerivatives 4.0 International License [1]. Since the material is not being used for commercial purposes, and rather for analysis in the hope to discover information about student behaviour, this project may use the data provided.

Data description Data was provided in 5 separate CSV files. File task_3_4 contains 6 columns and 346428 rows. Three of the columns (AnswserId, QuestionId and AnswerId) are keys that are present in the other files as columns, which makes it so that the files can be joined on these keys. Column IsCorrect contains binary data for the correctness of the answer (0 - Wrong answer, 1 - Correct answer). For each multiple-choice question, there were 4 possible answers (A,B,C,D) which are modelled as 1,2,3,4 in the columns CorrectAnswer and AnswerValue.

Other files contain more variables, but the ones considered in this study are:

1) Confidence, from Answer_Meta_Data_task_3_4, which contains 1508917 records. This has a value of 0 - 100, 0 not tbeing confident at all and 100 being fully confident in the answer. 2) DateAnswered from the same file, contains the date a question was answered 3) SubjectId from Question_metadata_task_3_4, which contains 948 records, and has the Ids of the subjects of the questions in 4 levels (each level is more specific) 4) DateOfBirth from Student_Meta_Data_Task_3_4, which contains 6148 records, and has the date of birth for the students.

Data processing The data set was created by joining task_3_4 with Answer_Meta_Data_task_3_4 on AnswerId, and joining this dataset with Question_metadata_task_3_4 on QuestionId. All records with no confidence reported were dropped, and all other variables apart from (QuestionId, AnswerId, IsCorrect, CorrectAnswer, AnswerValue, Confidence and SubjectId) were dropped, leaving 346428 records.

To find the age of a student at the time of answering a question, another dataset was created, creating a new variable (AgeAtQuestion) that subtracted the date of answering a question from the date of birth, dropping all records for students that did not have both of these values.

4 Exploration and analysis

Logistic regression of confidence on the correctness of an answer was plotted in Figure 1. Data from the Confidence and isCorrect variables were used as the training data for the model. After the model was fitted, the data was passed as the test data, and the model showed it had a precision of 66% in predicting whether an answer would be correct for certain confidence. Beta zero is found to be -0.84 and applying Equation 1

$$f(\beta_0) = 1/(1+exp(\beta_0)) \tag{1}$$

We conclude that when the confidence is zero, the probability of getting a correct answer is around 0.3. From Beta 1 (0.0184), once again applied to Equation 1, we deduce that for every percentage increase

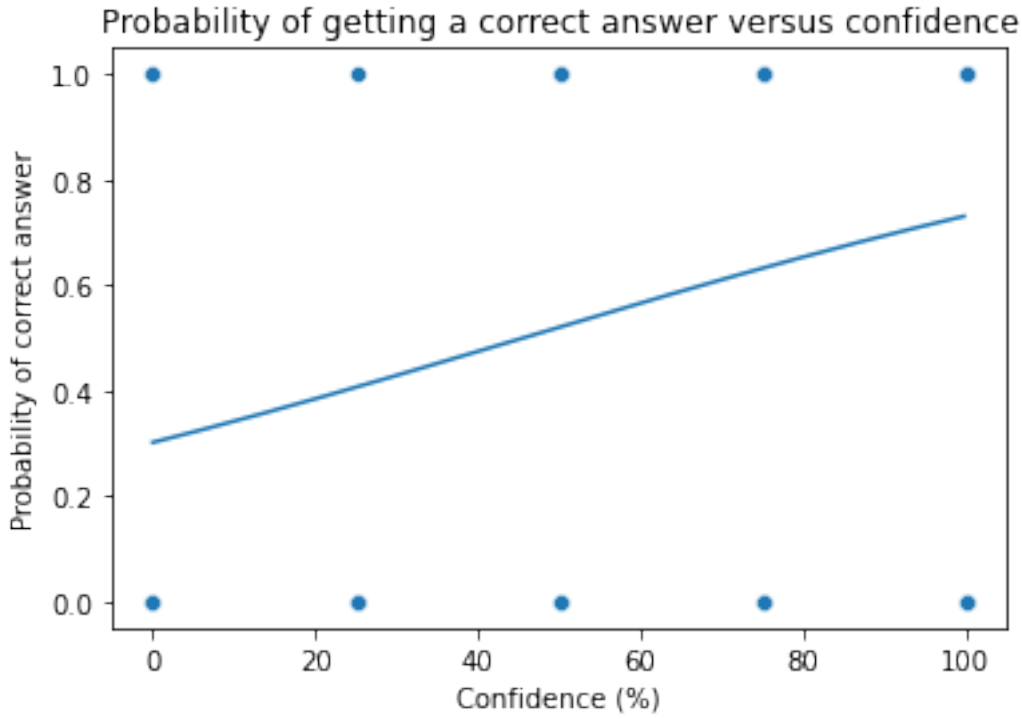

Figure 1: Logistic Regression of confidence percentage on probability of getting the answer correct.

in confidence, the probability of getting an answer correct increases by around 1.8%. This suggests that even though a student may have no confidence at all in their answer, they are still more likely to get the answer right than if it were to be selected at random (0.3 vs 0.25 probability).

A scatter plot was plotted for all the questions with the proportion of the total answers that were overconfident in Figure 2. As is clear from the graph, overconfidence is quite a common behaviour and suggests that there are miss conceptions in some questions. There were 99 questions above the threshold of 40% overconfident proportion, which implies that for those questions, more than 40% of the people had overconfident responses. These questions are worth analysing as they may bring up common misunderstandings among certain topics.

In Figure 3, The first level of subjects discussed in this study is looked at for the questions that had over 40% overconfident answers reported. Number and Geometry and Measure have many more questions than Algebra (Each with at least more than 3 times as many). This implies that the students in general have more misconceptions about the fields of Number and Geometry and measure than Algebra. Similar bar charts have been plotted for all of the levels of subjects (included in the notebook) but are not considered in the study due to constraints in the size of the document.

Nevertheless, in the overconfident questions, the question that showed this behaviour the most showed 86% overconfident responses, the answers were reported in Figure 4. This was only plotted for this specific question, as it was the one that stood out the most from the data set, as not only did more than 80% of the population get this wrong, but they thought that their answer was correct. This in particular shows a big misconception for the question, and it may be worth looking at why most of the students picked answer 1, when the correct answer was answered 4.

A similar scatter plot to Figure 2 was done for the underconfident answers, as shown in Figure 5, but it is clear that except in some particular cases, the underconfidence behaviour is much less usual than the overconfidence (this is to be expected to a certain degree due to there being 4 questions and 1 answer, so it is more likely to get an answer wrong than right).

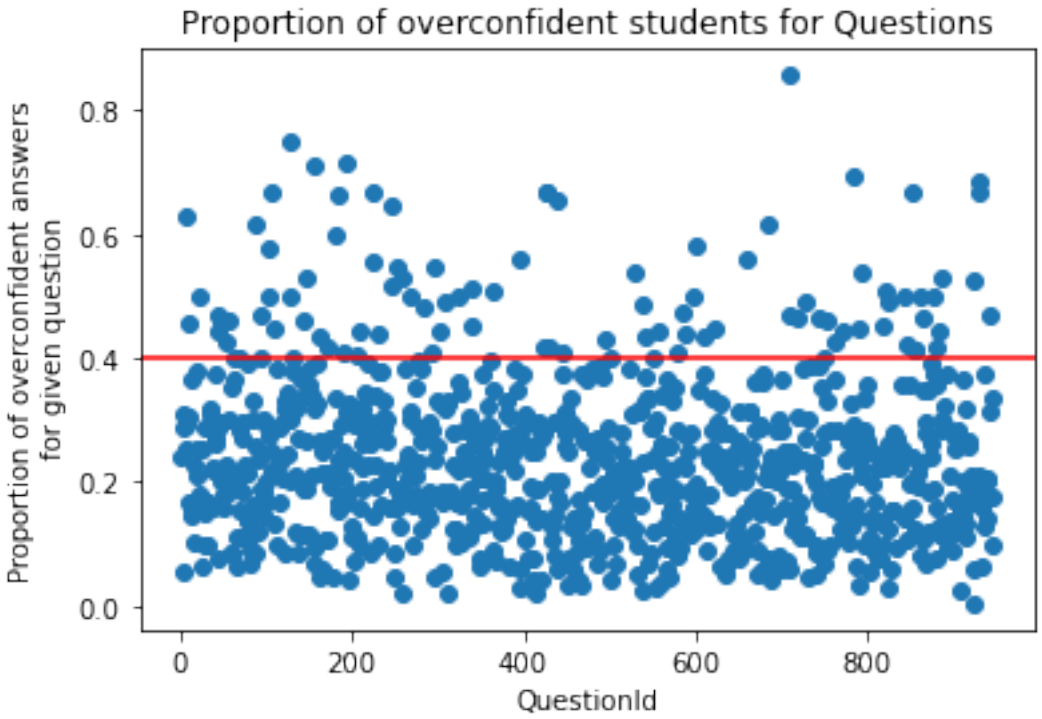

Figure 2: Scatter plot of the proportion of overconfident students (Reported a confidence > 50 and got the question wrong) from everyone that answered that question, for every question. The red threshold line at 0.4 is to mark the questions where more than 40% of the students were overconfident, and further analyse these questions.

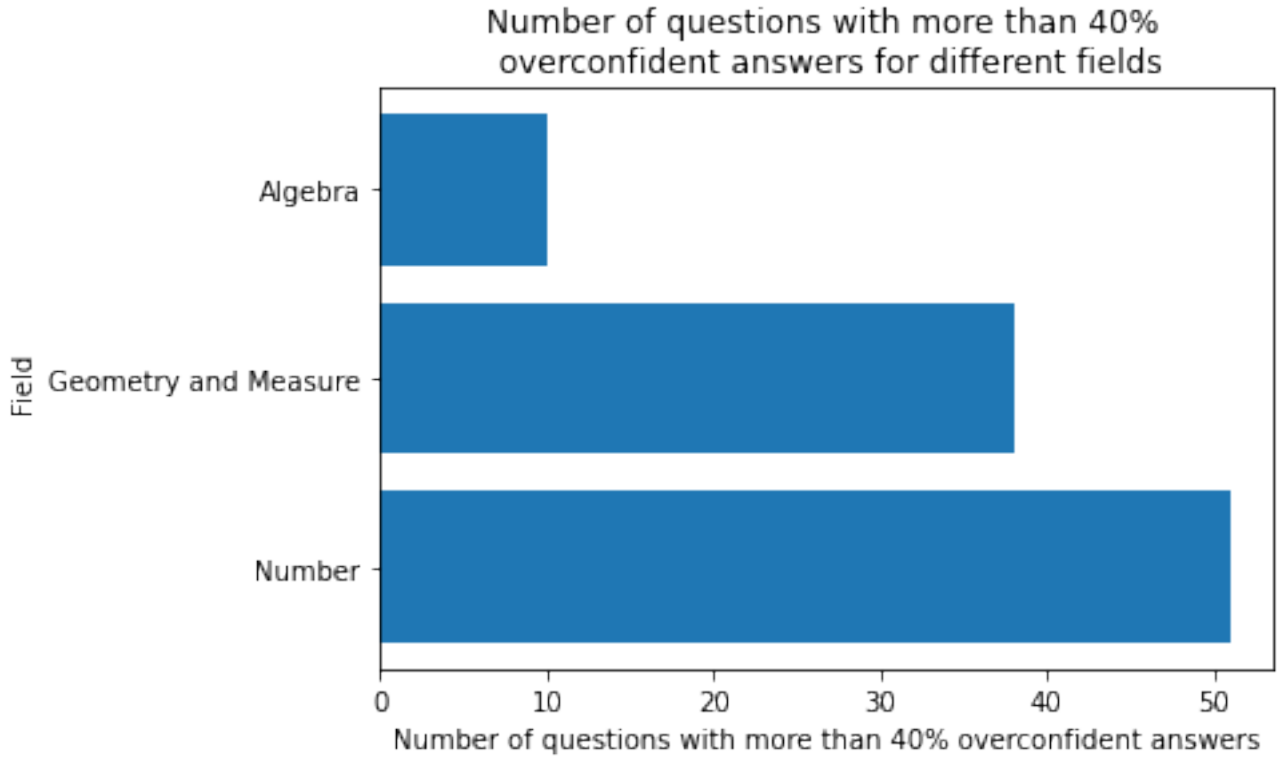

Figure 3: Bar chart of the number of questions which were answered with more than 40% overconfident answers for the different maths fields in the study.

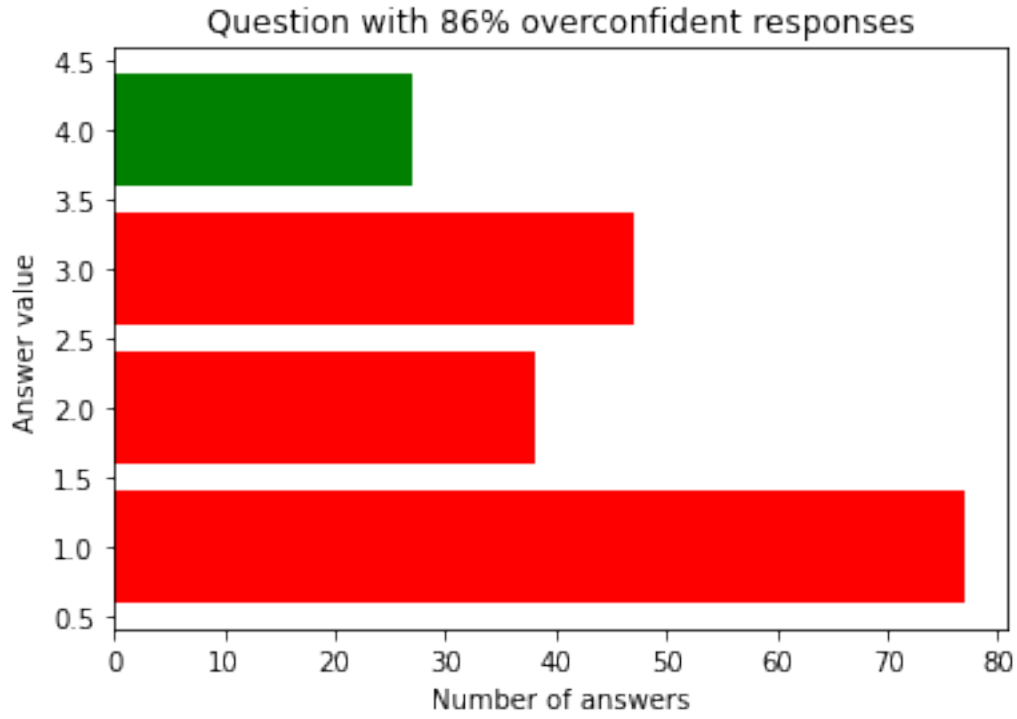

Figure 4: Bar chart of answers recorded for the question with most overconfident responses. In red are the wrong answers and in green is the correct answer.

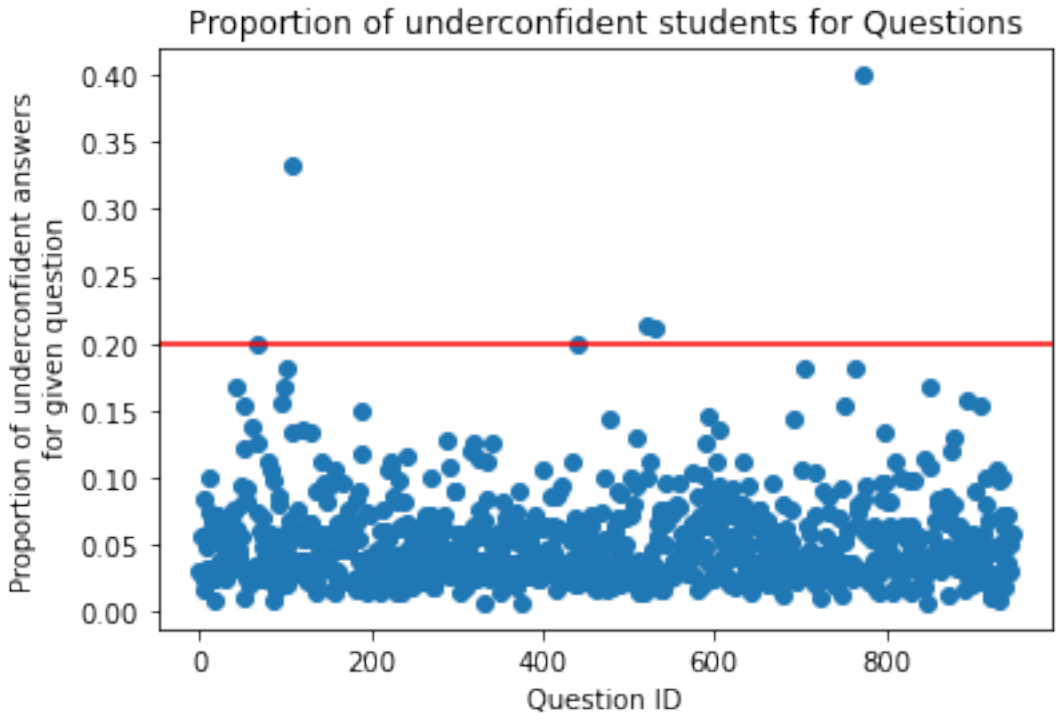

Figure 5: Scatter plot of the proportion of students that were underconfident (Reported a confidence < 50 and got the question right) from everyone that answered that question, for every question. The red threshold line at 0.2 is to mark the questions where more than 20% of the students were underconfident

5 Discussion and conclusions

Summary of findings From this study, some conclusions can be drawn: From the logistic regression, it was deduced that even though someone was not confident at all in their answer they were still better off than picking a random answer. This may suggest that people can to a certain extent deduce an answer for a multiple-choice question despite not knowing what the question is asking. It aslo showed that the more confident, the more likely they were to get a question correct.

Also, it was found that the overconfident behaviour was much more common than the underconfident behaviour (4.5 times more likely). This implies that there are more misunderstandings (when someone is overconfident, they think they have a correct answer, but it is wrong, which implies a misunderstanding in the knowledge), than uninformed answers (when someone is underconfident, they think they do not know the answer so they may pick a correct answer without knowing why it is correct).

Evaluation of own work: strengths and limitations Overall, the study does a good job of identifying student behaviour patterns with their reported confidence level, and how this relates to general misconceptions and in what fields students may find them.

It is however largely limited by the required size of the document, and a deeper analysis of the figures would be positive for the study. Also, some of the visualisations are not included due to this issue, and more of the variables provided could have been analysed to find if they had any effect on student behaviours.

Comparison with any other related work Similar findings were shown by Donald A Curtis et al. [2], as this article complies with the fact that students, in general, tended to have more misconceptions than being uninformed. It however contrasts the findings from Rafael Henrique Rangel et al. [3] which suggest the complete opposite, that there are more misunderstandings than misconceptions when looking at the confidence levels reported.

This could be due to the field of study and perhaps varies according to where the studies are conducted. For example in the case of Rafael Henrique et al.'s study [3], students may be prone to be underconfident due to the subject of neurology being a delicate one, and students may prefer to say they are unsure and get a correct answer rather than saying they are certain and getting a wrong answer.

Improvements and extensions To further analyse the students' behaviour, it may be worth looking at how particular students answered questions as individuals. Instead of just looking at overall levels of confidence, looking at confidence levels reported by individuals for multiple questions may show behavioural patterns in reporting confidence for those individuals. Other variables provided in the data set could be used to then analyse if they have any effect on the confidence reported. In the notebook, I provided a data set with the age of the students answering the questions and the average confidence reported. This could be analysed to find more patterns, and other variables can be considered, but are not in this study due to the length constraints.

References

[1] *CREATIVE COMMONS — ATTRIBUTION-NONCOMMERCIAL-NODERIVATIVES 4.0 INTERNATIONAL — CC BY-NC-ND 4.0*. 2022. URL: https://creativecommons.org/licenses/by-nc-nd/4.0/.

[2] Donald A Curtis et al. "Does student confidence on multiple-choice question assessments provide useful information?" In: *Medical Education* 47.6 (2013), pp. 578–584. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/medu.12147.

[3] Rafael Henrique Rangel et al. “Sure, or unsure? Measuring students’ confidence and the potential impact on patient safety in multiple-choice questions”. In: *Medical Teacher* 39.11 (2017), pp. 1189–1194. URL: https://doi.org/10.1080/0142159X.2017.1362103.

[4] Zichao Wang et al. “Diagnostic questions: The neurips 2020 education challenge”. In: *arXiv preprint arXiv:2007.12061* (2020).