Student learning on the EEdi educational platform

Sachin Loecher

April 11, 2022

1 Overview

The relationship between confidence and attainment is something most of us inherently understand, but recent research provides the evidence which reveals a causal relationship between attitude and mathematical performance [3], [5], [6].

This evidence is extensive, and highlights the importance of instilling a sense of confidence in students at an early age, as it can affect their performance and future career choices. Even more interesting are the demographic patterns found in the relationship between confidence and performance [4]. There is evidence that factors including gender, socioeconomic status, and age affect a student's confidence, and thus their performance. The interactions between these variables could explain observable patterns, and give us insight into the psychology of learning. Further exploring and understanding these patterns in new data could also provide new evidence based approaches to mathematical learning, and minimize demographic inequality.

2 Introduction

Context and motivation Mathematics serves as a filter in education, often resulting in high status and pay for successful students [2]. These minor differences in K-12 schooling result in differences in career choices, which exacerbates the under-representation of women in math-related fields and the pay gap. y Educational Data Science makes sense of the problems of educational systems, and helps us gain an understanding of how we can use evidence based approaches to find their solutions. Training educators in the field of Educational Data Science could help adapt educational systems for the needs of minority groups. Exploring the data could also provide answers to questions about learning psychology or demographic inequality

Previous work There exists previous work uncovering interesting results. It has been shown that a student's confidence is associated with their attainment, and their inclination to pursue a math-related career [7]. More interesting though are the patterns discovered in demographic variables such as socioeconomic status, gender, and age. It has been shown that confidence is generally lower amongst females and those socioeconomically disadvantages [4]. Previous research also shows that there exists a hyper-correlation which reveals that the higher a students confidence on an incorrect question, the higher the probability that they will answer the subsequent question correctly [1][3]. In the study[3], this effect was greater amongst females and younger students. Most older studies involved small data sets, but the accessibility of online learning tools has given us access to large data sets, allowing us to further explore and verify such effects.

Objectives We wish to explore how a students confidence when answering a question relates to their learning behaviour. We will show how each variable interacts with student confidence and performance, and whether there exist significant demographic differences in confidence and performance. This will hopefully either verify the previous research on new data, or contradict it. We will also observe biases in the data, such as selection bias in missing data, and the effects of sample size. Additionally we will explore patterns in individual student confidence subsets, such as when a question is incorrect.

The questions we would like answers to include: What effect does confidence have on a students attainment? What demographic features significantly impact student confidence and attainment? Are there interactions between these features?

3 Data

Data provenance Who created the data set(s)? How you have obtained it (e.g., file or web scraping), and do the T&Cs allow you to use obtain the data for the project?

The data set used was originally published as part of the NeurIPS 2020 Education Challenge, and was provided as a file by the online educational platform EEdi [8].

All participants consented for their anonymous data to be used. We use data collected under the lawful basis of legitimate interests (GDPR Art. 6 (1) (f)). The terms and conditions allow us to use the data non-commercially.

The data was collected through EEdi diagnostic questions for students from primary to high school. A diagnostic question is a multiple-choice mathematics question with 4 possible answer choices, one of which is correct. Additionally, students recorded their confidence (0,25,50,75,100) for each question. 5 data sets were provided, one with the response data, one with the student data, one with the question data, one with the subject data, and one with the answer data. After merging and matching the data, we have a data set where each response includes the User Id, the Question Id, the Answer Id, whether the answer was correct or not, the correct answer, the given answer, the date answered, the students confidence in their response, the Group Id for different classes, the Quiz Id for different quizzes, the students gender, whether they are receiving income support or not, and their date of birth. This data consists of 1,382,727 responses to 948 questions by 4918 students.

Data description

Data processing We created the data set we are using by merging the response data with the answer data, and then merging that with the student data. That way we ended up with one data frame with each response and the information about the answer and the user. Later we also grouped by the User Id, in order to get summary statistics for each user such as their average confidence.

This data set had many missing values for student confidence, whether a student was receiving income support or not, and the student's date of birth. To explore the potential selection bias in this opportunistic data set, binary indicator variables such as 'Missing Confidence' were added and explored later. The fitted models worked on cleaned data, in which the points with missing information were removed.

4 Exploration and analysis

We first conducted Exploratory Data Analysis by exploring and visualizing most of the feature distributions, and how they related to confidence and proportion of correct answers. In this paper we chose to aggregate over User Ids and Question Ids since these were the most relevant and interesting distributions.

We first visualized the user and question distributions by average confidence and proportion of correct answers. As we expected, there is much more variance in the User distribution than in the Question distribution, because the sample size for each question is much larger than that of each user. We also attempted to explain the outliers by incorporating Group Id (Class), but found no relationship between the best performing students and their class.

We then visualized the average confidence and proportion of correct answers distributions by age group, and found there to be no pattern. This could be because of the data itself and that almost all of the users were the same age.

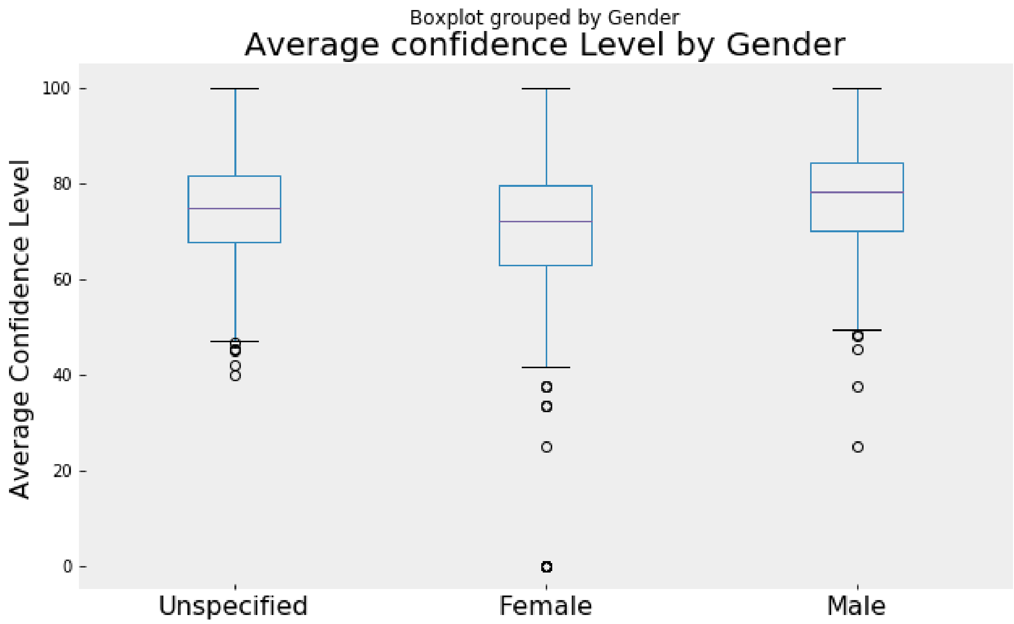

Figure 1: Average Confidence by Gender: The t-tests indicate that Male Confidence is significantly higher than Female Confidence (p = 0.00001187)

We then explored the User frequency distribution to see the distribution of sample sizes. We found that a large portion of the students answered very few questions. Interestingly we also see that this distribution is slightly bi modal in nature, with another peak at about 550 questions. The large variance seen in the distribution points us towards using a model which accounts for sample size later on.

We then plotted the distribution of confidence levels, and the proportion of correct answers by confidence level. There is a clear gap between confidence and attainment, indicating that these students are overconfident.

Next we explored the demographic effects of gender. We made box plots of the average confidence and proportion of correct answers by gender, and then computed t-tests to determine if the difference between male and female was significant (FIGURE 1). The t-tests indicate that there was a significant difference between male confidence and female confidence ($p = 0.00001187$), but no significant difference between male attainment and female attainment ($p = 0.6737$).

We then explored whether missing confidence had anything to do with selection bias, and if there is a difference between the average correctness for those who reported confidence and those who didn't (FIGURE **??**). The t-tests indicated a very large and significant difference, which indicates that students were much more likely to report confidence when they were more confident.

The next step was to see if a more confident students performed better. We can see in the box plots, that proportion of correct answers goes up with each confidence level, which is expected (FIGURE 2).

We then explored the demographic effect of income support. We created box plots and ran t-tests to determine whether socioeconomic status had a significant effect on the performance and confidence of students. As we can see, both confidence ($p = 0.00071$) and attainment ($p = 1.6426e^{-}9$) were significantly less in students with income support.

We then plotted confidence when incorrect by confidence when correct. As we can see, most of the points lie above the 45 degree line, so students were in fact more confident when answering correctly. The correlation ($c = 0.6286$) also shows that overconfident students were overconfident both when correct and incorrect.

We can also see the correlation between correctness and confidence by sample size (FIGURE 3). Here again, we see that more confident students do indeed score higher.

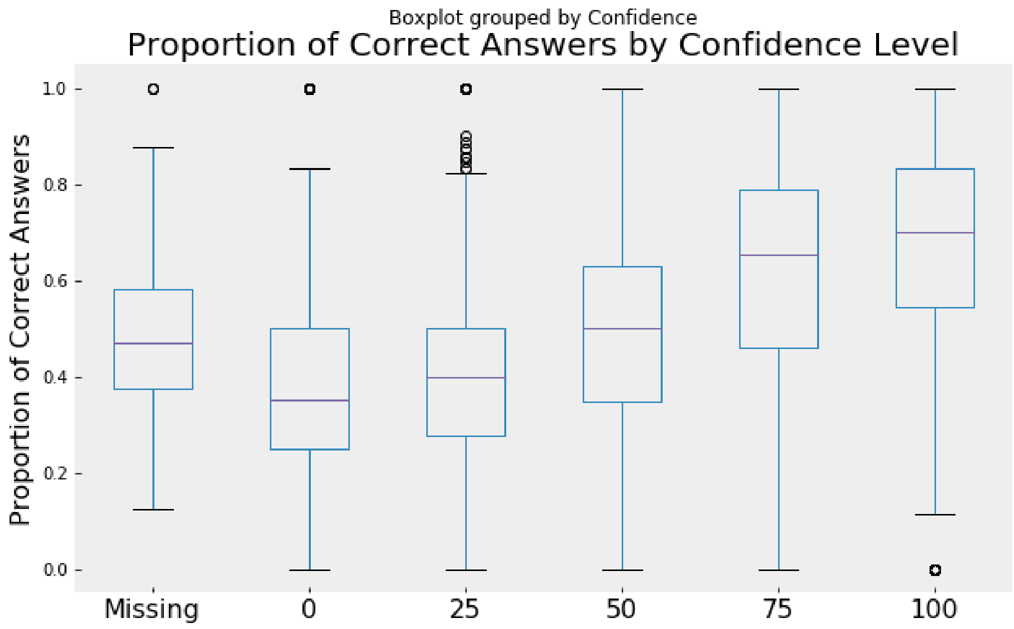

Figure 2: Proportion of Correct Answers/Ability by Confidence Level. We see a clear step wise increase in ability for each increase in confidence.

We plotted a scatter plot of age vs confidence as well, and found found no large association between age and confidence.

We then tested the claim of the paper[3] that the confidence gap between genders increases with income support. We found no significant difference in confidence between the genders receiving income support, which contradicts this.

The first model we fitted was a very simple ordinary least squares linear regression model to predict a students average confidence by Income support, Gender, Proportion of Correct Answers, and Age. The only two significant features in this regression were Proportion of Correct Answers and Gender, as one can see in table 4.

	coef	**std err**	**t**	**P> |t|**
Intercept	23.8105	21.489	1.108	0.268
PremiumPupil	-0.8218	1.006	-0.817	0.414
Age	1.2920	1.515	0.853	0.394
isMale	4.5800	0.900	5.088	0.000
AverageCorrect	42.0704	1.954	21.527	0.000

Table 1: Summary table from OLS Regression

The second model we fitted was a weighted least squares linear regression. We decided to use this model, because it accounts for the lack of homoscedasticity. This time, Proportion of Correct Answers, Gender, and Age were all significant features as one can see in the summary table 3.

The last model we fitted was an ordinary least squares linear regression model with all 2-way interactions. We used this because we wanted to explore the significant interactions between the features. The only significant interaction was between Proportion of Correct Answers and Income Support as one can see in the table 3 ($alpha = 0.05$).

The methods we used to determine the importance of each feature included t-tests to compute significance without the normal assumption, ordinary least square regression, weighted least square

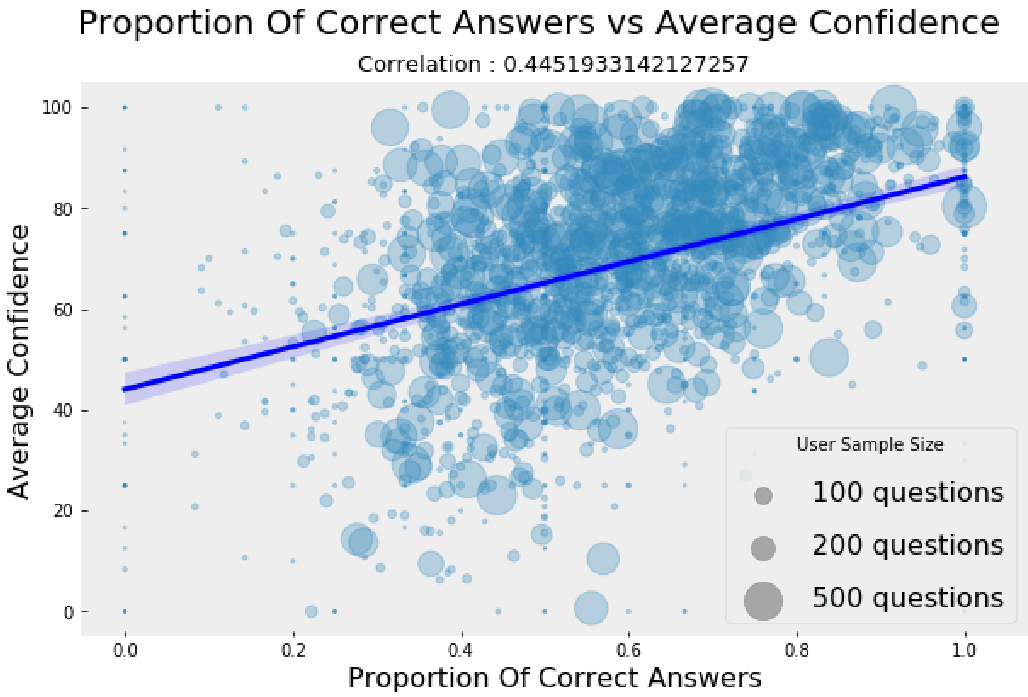

Figure 3: Proportion of Correct Answers vs Average Confidence. We see a strong correlation, meaning the students with higher performance generally have higher confidence as well. We also notice the abundance of very small sample sizes, which will need to be considered in our models.

	coef	std err	t	P> |t|
PremiumPupil	-0.6114	0.919	-0.665	0.506
Age	2.7806	0.108	25.689	0.000
isMale	3.7063	0.753	4.923	0.000
AverageCorrect	52.9197	2.306	22.954	0.000

Table 2: Summary table from OLS Regression with 2-way interactions

linear regression, and ordinary least square linear regression with 2-way interactions. When we use "Proportion of Correct Answers” as a proxy measure for students’ ability (sometimes referred to as “facility”), we find a strong and significant relationship between this ability and the mean confidence for each student. The findings of the models and t-tests also show Gender to be a major factor in confidence. When we use Weighted least squares, Age becomes a significant factor as well. The interactions model’s findings indicate that the general overconfidence of the students seems to decrease with increasing ability ("Proportion of Correct Answers").

5 Discussion and conclusions

Summary of findings We found large variations in students self reported confidence as well as correctly answered questions at every aggregation level: Group Id, student Id, Quiz Id, etc.. We found a strong and significant relationship between a students ability ("Proportion of correct answers") and their average confidence. We also found a significant relationship between Gender and Average Confidence. Age was also considered a significant factor in Average Confidence, when accounting for sample sizes. Generally, students seem to be overconfident, meaning that the actually achieved percent of correctly answered questions is much lower than the self-assessed confidence. This observed overconfidence is

	coef	std err	t	P> |t|
Intercept	39.3368	64.354	0.611	0.541
AverageCorrect	29.5848	94.976	0.311	0.755
Age	0.0424	4.542	0.009	0.993
isMale	20.8860	43.053	0.485	0.628
PremiumPupil	-50.3676	48.481	-1.039	0.299
AverageCorrect:Age	1.0924	6.705	0.163	0.871
AverageCorrect:isMale	-0.4846	3.917	-0.124	0.902
AverageCorrect:PremiumPupil	-9.3433	4.358	-2.144	0.032
Age:isMale	-1.0707	3.033	-0.353	0.724
Age:PremiumPupil	3.9559	3.417	1.158	0.247
isMale:PremiumPupil	-2.9145	2.013	-1.448	0.148

Table 3: Summary table from OLS Regression with 2-way interactions

more pronounced for Males than for Females, and seems to decrease with increasing ability as shown in our extended linear model which includes interaction terms.

Evaluation of own work: strengths and limitations Wile we have to assume that values are "missing at random", there is evidence for violation of this assumption, as questions are answered less correctly on average when confidence is missing. Unfortunately, the majority of questions did not contain a value for confidence, so any question pertaining to confidence could be tackled only with a reduced sample size We did not think Ordinary least squares regression was a very appropriate model to use because of the greatly varying sample sizes per student, which leads to a violation of the assumption of homoscedasticity. Clearly, the measurement is much less precise for low sample sizes. One solution was the Weighted least squares regression, where we use sample sizes as weights, as sampling variance decays inversely proportional with sample size.

The data used was also collected online, and may potentially not be generalised to the classroom. Another limitation is that confidence is a self-reported measure, and could have unknown human biases in itself such as exaggeration.

Although there were many missing values for Confidence, Age, and Income Support, the rest of the data was large enough and well collected. and the results we have found are significant and reliable.

Comparison with any other related work Most of the results we have found verifies most previous research [3], [4][6], and the claim that confidence is associated with higher performance. We also verify the effect of certain demographic variables like Gender and Income Support. However, the claim that confidence decreases with Age was contradicted in our study, although this may because of the lack of quality of our Age data.

Improvements and extensions In future research, a major improvement would be better quality data. Although this data set was a great size, it was opportunistic and thus contained biases and lots of missing information. Another solution to the low and varying sample sizes we looked into was Empirical Bayes. In the code provided, we have provided an example of how this would be used on our data.

Another interesting approach to finding trends in the data would be to use a decision tree. This could provide insight into the structure of the data, and lead to better models in the future. We have also included an example of how a decision tree could be applied to our data in the code provided.

There are also loads of other models and tests that could be looked into, such as the non-parametric Wilcoxon tests, which allows for varying sample size as well.

Future research could also look into these effects over time, and find trends in learning behaviours and demographic variables.

References

[1] B Butterfield. "Errors committed with high confidence are hypercorrected". In: *Journal of Experimental Psychology: Learning, Memory, and Cognition* (2001). Retrieved on 10 April 2022. URL: `https://doi.org/10.1037/0278-7393.27.6.1491`.

[2] Patricia Campbell. "So what do we do with the poor, non-white female? Issues of gender, race, and social class in mathematics and equity patterns and reciprocal relations". In: *Peabody Journal of Education* (1991). Retrieved on 10 April 2022. URL: `https://doi.org/10.1080/01619568909538640`.

[3] Colin Foster. "Confidence and competence with mathematical procedures". In: *Educ Stud Math* (2021). Retrieved on 10 April 2022. URL: `https://rdcu.be/cK31S`.

[4] C Ganley et al. "Mathematics confidence, interest, and performance: Examining gender patterns and reciprocal relations". In: *Learning and Individual Differences* (2016). Retrieved on 10 April 2022. URL: `https://doi.org/10.1016/j.lindif.2016.01.002`.

[5] Kishor N Ma X. "Graphs in statistical analysis". In: *Journal for Research in Mathematics Education* 28.1 (1997). Retrieved on 10 April 2022, pp. 26–47.

[6] I Mullis et al. "TIMSS 2019 International Results in Mathematics and Science". In: *TIMSS PIRLS International Study Center* (2020). Retrieved on 10 April 2022. URL: `https://timssandpirls.bc.edu/timss2019/international-results/`.

[7] Alisa Paulsen. "Basic Confidence Predictors of Career Decision-Making Self-Efficacy". In: *The Career Development Quarterly* (2011). Retrieved on 10 April 2022. URL: `https://doi.org/10.1002/j.2161-0045.2004.tb00951.x`.

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