DEVELOPING REGRESSION HEURISTICS THROUGH ARGUMENTATION IN COMPUTING LABORATORY

Ken W. Li
Retired Scholar
dr_ken_li@yahoo.com.hk

Developing regression heuristics for students is not straightforward. It would be better to organize class settings to facilitate small group discussions; they can talk about how to address a question of common concern through argumentation such that their thoughts would be articulated. The students in the present study were divided into small groups in a computing laboratory in order to increase opportunities for peer discussion. An observation study was then conducted; the contents of their discussions were analyzed. It was found that argumentation was characterized by the question-and-answer exchanges that shaped the flow of regression tasks on which they worked. They were comfortable in presenting their standpoints or responses, and the responses were taken seriously by their peers. They raised questions when they found peers' feedback vague. Besides, group interaction was associated with positive responses, thus articulating students' thoughts, enriching thinking context, and broadening thinking perspective.

INTRODUCTION

Statistical heuristics are tools for inductive reasoning when making judgments or decisions (Nisbett et al., 1983). Proficiency in applying statistical heuristics may be developed through classroom argumentation based on Kuhn (1992) who argued that argumentation was a social activity through which students interacted among themselves or with their teacher. They made their ideas available via communication to others for comment, suggestion, and argument such that their thoughts were articulated and thinking perspective was broadened, thus improving human reasoning skills in accordance with Schwarz and Asterhan (2010) as well as Wegerif (2015). To foster argumentation, it would be better to organize classroom settings to facilitate small group discussions (Gillies, 2012) conducive to collaborative interactions that is somewhat known as collaborative argumentation-based learning (Noroozi et al., 2012). Nowadays, collaborative learning commonly takes place in computer-supported environments that would also facilitate peer argumentation (Asterhan, 2012).

EMPIRICAL STUDY

An empirical study was set within a technology-enriched classroom, i.e., a lecture theatre and a computing laboratory were equipped with computers connected to the internet and the school intranet, computer monitors, computer software, audiovisual system, and so on. Technology is not confined to these physical artefacts but also refers to cultural resources that would facilitate social interactions (Wegerif, 2015) and foster collaborative learning (Schwarz & Asterhan, 2010).

Fifty-eight higher diploma students (aged 19-22) enrolling in the second year of the three-year ASC (Applied Statistics and Computing) program voluntarily participated in the study. The program aimed at equipping students with statistical and computing knowledge together with practical skills. The ASC graduates would find employment as statistical officers, research assistants, etc. The students were divided into small groups so as to increase opportunities for peer argumentation. Each group of students shared the same computer and monitor and took turns in programming Excel as they worked together on the learning tasks in a practice session conducted in a computing laboratory.

In the practice session, the students were asked to download a set of secondary data consisting of six variables (i.e., y, x_1, x_2, x_3, x_4, x_5) from the school intranet. The data context was about social security payment. Among these six variables, y (the amount of social security spent), x_2 (the number of ill health cases), and x_4 (the number of unemployed cases) were selected for them to practise regression heuristics by completing the following four tasks in a laboratory worksheet that was provided by the teacher: 1) to build two simple linear regression models using Excel; 2) to evaluate the statistical significance of regression parameters using hypothesis testing; 3) to give the regression models after the above testing; and 4) to discuss which one of the two regression models best describing how social security budget was spent. The fourth task is about to deploy regression heuristics in such a way as to evaluate the goodness of fit among regression models by means of

In: R Helenius, E Falck (Eds.), *Statistics Education in the Era of Data Science*
Proceedings of the Satellite conference of the International Association for Statistical Education (IASE),
Aug-Sept 2021, Online conference. iase-web.org/Conference_Proceedings.php

residual plots, the coefficient of determination (R^2), and the significance testing of overall model fitting or regression slopes. The evaluations are grounds for building an optimal model or clues in a model refinement. After completing the tasks, they presented their works on the laboratory worksheet.

When the students were attempting the tasks collaboratively with their peers, there was a substantial amount of students' verbal exchanges serving different purposes. For instance, students who respond to their peers positively without critically evaluating what they are told use cumulative talk. The talk is mainly for attempting low-collaborative tasks or maintaining social interactions. Exploratory talk features very largely in high-order of thinking mandatory for students to develop regression heuristics; they critically evaluate what they are told prior to accepting. Perhaps, disputational talk is used to challenge someone's proposal merely based on their personal point of view. These three types of talk originate from Mercer (1995); both cumulative and exploratory talks are conducive to peer argumentation when developing regression heuristics, whereas disputational talk is not constructive.

The verbal exchanges were audio-recorded and transcribed in full, with relevant excerpts being selected for analysis using Mercer's framework (1995) in order to address the research question of how regression heuristics would be developed through peer argumentation in the technology-enriched classroom. Unfortunately, most student participants appeared to be anxious at the prospect of having their dialogues audio-recorded, and as a result talked less frequently and less audibly. The conversation could not be used for analysis, eventually ending up with the recording of the conversations of only one group of three students, identified by codes, D, K, and P. The following analysis thus uses excerpts from their conversations.

ANALYSIS AND DISCUSSIONS

The students studied the given set of data in terms of its context, measurements, and measurement units after downloading the data embedded in an Excel spreadsheet from the school intranet. Shortly after, student K ran Excel programs to build regression models, the regression results displayed on their computer screen. K like other students were proficient at building regression models using Excel because they had accomplished similar tasks in the past few practice sessions. The three students, D, K, and P now moved on to develop regression heuristics that was one of the last few topics in the entire process of simple linear regression modelling.

Excerpt

1.	K:	The result is here.
2.	D & P:	So quick! I hadn't read all the questions.
3.	P:	What do we compute?
4.	K:	To find the $\hat{\beta}_0$ and $\hat{\beta}_1$
5.	P:	Do we just do the first one?
6.	D:	We need to attempt all the questions.
7.	D & K:	Write the answer.
8.	K:	$\hat{\beta}_0$ equals to this figure.

K's prompt Excel programming actions and announcement, "*The result is here*" contrasted with D's and P's slowness. P asked what to compute and her question influenced the flow of regression modelling. K replied, "*To find the* $\hat{\beta}_0$ and $\hat{\beta}_1$". P understood what to do but asked a question irrelevant to the previous issue or regression tasks, "*Do we just do the first one?*" To maintain interaction among themselves, D replied, "*We need to attempt all the questions*", they all agreed. D and K said to write the answer when K was pointing at the value of $\hat{\beta}_0$ displayed on their computer screen. K's non-verbal cue further clarified P's query.

According to Mercer's categories (1995), Excerpts 1-8 are cumulative talk through which the students responded to their peers positively. They accepted but did not critically evaluate what was told probably because the tasks were so far straightforward. Both K and D made contributions in that

K programmed Excel and D outlined the initial steps of regression heuristics; whereas P obtained clear directions of what to do.

Excerpt

9.	P:	-49946. How many places of decimal do we need?
10.	K:	Four places of decimal
11.	P:	Four places of decimal
12.	D:	Two places of decimal are OK.
13.	K:	Four places of decimal must be sufficient.
14.	P:	Four places of decimal must be sufficient.

Student P read the value of $\hat{\beta}_0$ aloud when she was putting it on the laboratory worksheet. She raised an issue of reporting the number of decimal places on the laboratory worksheet. The three students, P, D, and K then negotiated the accuracy of regression results for reporting when encountering an ambiguity, i.e., model accuracy versus model practicality. The students had learnt that, ideally, parameters of a regression model possessing more decimal places can make more accurate prediction but whether such accuracy was a top concern is in doubt when taking model practicality into consideration. Instead, regression parameters with too many decimal places might create an interpretative burden and computation problems far more than gaining a non-significant accuracy. As such, both model practicality and model accuracy are also critical issues in regression heuristics. Excerpts 9-14 are exploratory talk in nature because their negotiation indicated that they were in different positions in evaluating their regression model as to whether or not too many decimal places were negligible when contrasting model practicality with model accuracy. A compromise agreement was at last arrived at using four decimal places.

Excerpt

15.	Teacher:	You've used the Regression Analysis Tool. You're going to test whether or not $\hat{\beta}_0$ is equal to zero; $\hat{\beta}_1$ is equal to zero. Should you do two separate statistical tests?
16.	D, P & K:	Two separate statistical tests
17.	Teacher:	Two separate (statistical tests)
18.	D, P & K:	OK.

While the students were discussing how to evaluate the statistical significance of regression parameters, β_0 and β_1 using hypothesis testing, the teacher intervened to check their learning progress. He found the students could not manage the evaluation task even though they could use the Regression Analysis Tool in Excel to yield statistical output. Prior to reading the output, it was necessary to know which statistical tests (F-test or t-test) and how many statistical tests ought to be conducted. He hinted that they might evaluate the statistical significance of each individual regression parameter in two separate tests. He then raised a question so as to promote peer discussions. The three students had a quick discussion and responded to show their awareness of these statistical routines when evaluating the significance of regression parameters. The teacher recapped their response and they confirmed, *"OK"*.

Excerpt

19.	Teacher:	How many programs do we need to run for these two models? … We want to look X_2 and X_4 separately, so how many programs do we need to run?
20.	D:	Uh, …
21.	Teacher:	For a simple linear regression model, … how many? Do you understand what I mean?
22.	D:	I know what you mean but does it include the program (we have just run)?
23.	Teacher:	I just want to look at the relationship. I am talking about the relationship, (that) means the simple linear regression model.
24.	D:	Huh!

25.	Teacher:	We have looked at the correlation already. We proceed to looking at the exact relationship. The "exact" means, how can we put y equals to a value for $\hat{\beta}_0$ and a value for $\hat{\beta}_1$ times *x*?
26.	D, P & K:	Huh! Huh!
27.	Teacher:	How many programs do we need to run for two separate models?
28.	P:	Two
29.	Teacher:	One? Two? Three?
30.	D, P & K:	Two
31.	Teacher:	Two. Only two.

As the students seemed inert, the teacher used questioning extensively to structure their thinking. The teacher raised a question in line with the argument he developed in Excerpt 19. First, the teacher wanted the students to concentrate on building two simple linear regression models rather than also considering a multiple regression model on account of an interaction effect, $X_2 X_4$. Second, the teacher also asked students to identify programming requirements and translate them into Excel programming prior to implementation because proper Excel programming would facilitate regression heuristics. While the students were thinking about what to answer; the teacher re-phrased his question, "*We want to look X_2 and X_4 separately, so how many programs do we need to run?*" They did not answer the question directly because they were not sure whether the number of Excel programs to be run should include or exclude the one they already had when answering the question. D thus asked, "*Does it include the program (we have just run)?*" to clarify this issue confronting them. The teacher did not respond to her question directly but emphasised that studying the data relationship extended to knowing how *y* was affected by *x* beyond merely knowing the degree of linear association between *y* and *x*. The teacher decided that P answered, "*Two (models)*" without confidence so that he offered three choices of answers, "*One? Two? Three?*" After a quiet discussion, they all chose, *"Two"* as their answer. In so doing, students' ability was assured beyond a mastery of Excel syntax. They were attentive to the teacher's scaffolding assistance.

Excerpt

32.	Teacher:	You did the first one, Y against X_2 and the next one is?
33.	D, P, K & Teacher:	Y *against* X_4
34.	Teacher:	Good! Very good! Yes, attach the output and test the significance. You just look at the output and discuss whether $\hat{\beta}_0$ under the model, X_2 is significant, OK?
35.	D, P & K:	Huh! Huh!
36.	Teacher:	… and also for X_4.
37.	D, P & K:	Huh! Huh!
38.	Teacher:	(The teacher pointed to Question 3), i.e., for Question 3, you want to look at $\hat{\beta}_1$.
39.	D, P & K:	Huh! Huh!
40.	Teacher:	OK? For X_2 model and …
41.	D, P & K:	OK!
42.	Teacher:	Individually
43.	D, P & K:	Huh!
44.	Teacher:	And Question 4, we want to construct the model. Again, "Construct" means we want to know the exact relationship between *y* and *x*. So we want to substitute the estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$ into the proposed model, OK?

45.	D, P & K:	Huh! Huh!
46.	Teacher:	This is the model we have constructed, OK? (The teacher was pointing at the regression model they had built.)
47.	D, P & K:	Huh!

The teacher recapped "*You did the first one, Y against X_2, …*", i.e., K had programmed Excel for building the first regression model, $Y=\beta_0+\beta_1X_2+\varepsilon$ and followed by questioning, "*…and the next one is?*" The question was posed by the teacher to recap the regression model the three students had already built, as hinting at the next modelling task. They were not confident about what to answer but they could complete the answer correctly after prompting by the teacher (Excerpt 33). The teacher gave praise, "*Good! Very good!*" to reinforce their answer (Excerpt 34). The students seemed threatened by being over-questioned but maintained an active dialogue with the teacher for social exchanges. The teacher eventually preferred to give explicit instructions, Excerpt 44 and Excerpt 46 with a non-verbal cue as a support for subsequent learning tasks. After the students had gradually picked up the hint, he left them alone to accomplish the remaining tasks to give them the opportunity to develop more autonomy and creative ideas in model building. The teacher's assistance here took the forms of questioning initially and instructing later in the hope that the teacher's voice would become the students' self-instructing voice to reorganize their thoughts.

Excerpt

49.	Teacher:	You have these two models. We want to find out a better model so we are going to look at the R^2. Look at the R^2 and then compare which one is better.
50.	D, P & K:	Huh!
51.	Teacher:	I will come back after you finish Question 4.
52.	D, P & K:	OK! Thank you.
53.	Teacher:	Try to think how to interpret your output; this is very important.

After the teacher had instructed the students to evaluate the fittings of two regression models using R^2, he did not offer further scaffolding assistance because the students saw how R^2 fitted in with the rest of the evaluation tasks. They then discussed how to attempt the fourth task about which one of the two regression models best describing how social security budget was spent.

On the laboratory worksheet the students submitted for evaluation, they showed how to build and justify the best regression model using regression heuristics. Building such 'best' regression models cannot merely be reliant on straightforward statistical procedures without involving regression heuristics in achieving the ultimate goal of regression modelling, i.e., arriving at a model that might serve our purposes as well as telling the truth (McLaughlin, 2001).

CONCLUSION

The development of regression heuristics is evident from argumentation that is characterized by a collaborative interaction among peers in the way that the students talked about how to address a question of common concern based on their own understandings, opinions, judgments or perspectives through verbal exchanges. They were comfortable in presenting their standpoints or responses, and their responses were taken seriously by their peers. They raised questions when they found peers' feedback unclear or ambiguous. These question-and-answer exchanges shaped the flow of regression modelling tasks on which they worked together.

There was also evidence that group interaction was associated with positive affective responses as well as exploratory and cumulative talks. Specifically, the students focused on about what to do and how to do in order to make joint decisions that led to a modelling progress using exploratory talk. On the other hand, cumulative talk was observed when the students proposed ideas or accepted the ideas of their peers without finding it necessary to give or seek justification. Both types of talk were valuable for knowledge construction. The verbal exchanges in exploratory talk inspired the students, to a great extent, would formulate and regulate strategies for model building. Apart from talk being used for addressing cognitive need, cumulative talk was also beneficial in building social relationships and fostering rapport between students that was a necessary condition for

carrying out critical discussions. As such, argumentation would be productive as long as social relationship is attained.

This study also shows that key roles played by a teacher in orchestrating social interaction between students in an IT environment that were aimed at developing regression heuristics. The teacher made decisions about when to supplement students' knowledge as well as skills and when to use questioning to stimulate thinking about Excel programming requirements; direct actions to check statistical significance of regression parameters; and promote intellectual exchanges between students to sort out the best regression model using R^2. To situate learning within students' capabilities, the teacher encouraged them to work on their own in order to regulate their strategies based on their own creation and interpretation of a regression model they found was the best. Unfortunately, some of their verbal exchanges might not be heard in the audio-recording owing to soft spoken voice.

Although the sample of dialogue analysed in this study do not include all students participating in the study, it is representative of the kind of talk observed throughout the observation period. It is also worth noting that the teacher found arguments written by all groups of students on their laboratory worksheets were grounded.

From the above, the implication is that learning tasks should be designed to encourage peer collaboration rather than independent work. Teachers should monitor their progress in argumentation in terms of its content and logic flow. Through argumentation, students' thoughts were articulated and regression heuristics together with concepts would become more refined.

ACKNOWLEDGEMENT

The author would like to express his sincere thanks to Dr. Barbara Jack for proofreading this article.

REFERENCES

Asterhan, C.S.C. (2012). Facilitating classroom argumentation with technology. In R.M. Gillies (ed.) *Pedagogy: New Developments in the Learning Sciences* (pp. 105-130). New York: Nova Science Publishers.

Gillies, R.M. (2012). Promoting reasoning, problem-solving and argumentation during small group discussions. In R.M. Gillies (ed.) *Pedagogy: New Developments in the Learning Sciences* (pp. 131-150). New York: Nova Science Publishers.

Kuhn, D. (1992). Thinking as argument. *Harvard Educational Review*, *62*(2), 155-179.

McLaughlin, M.P. (2001). *A tutorial on mathematical modelling*. Online: http://www.causaScientia.org/math_stat/Tutorial.pdf.

Mercer, N. (1995). *The guided construction of knowledge: Talk amongst teachers and learners*. Clevedon: Multilingual Matters Limited.

Nisbett, R.E., Krantz, D.H., Christopher, J., & Ziva, K. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90*(4), 339-363.

Noroozi, O., Weinberger, A., Biemans, H.J.A., Mulder, M., & Chizari, M. (2012). Argumentation-based computer supported collaborative learning: A synthesis of 15 years of research. *Educational Research Review*, *7*(2), 79-106.

Schwarz, B.B., & Asterhan, C.S.C. (2010). Argumentation and reasoning. In K. Littleton, C. Wood, & J.K. Staarman (eds.) *International Handbook of Psychology in Education* (pp. 137-176). Bingley: Emerald Group Publishing Limited.

Wegerif, R. (2015). Teaching and teaching thinking: Why a dialogic approach is needed for the twenty-first century. In R. Wegerif, R. L. Li, & J.C. Kaufma (eds.) *The Routledge International Handbook of Research on Teaching Thinking* (pp. 427-440). New York: Routledge.