An Empirical Study on Peer Discussion About Statistical Evidence in Computing Laboratory

Ken W. Li and Merrilyn Goos

Abstract To develop students' ability to justify or construct statistical evidence, students were placed into small groups in order to increase opportunities for peer discussions and social interactions in the classroom and beyond. An observation study was thus conducted to address the question of how peer discussion and student–teacher interaction would have a beneficial influence on learning the topic of statistical evidence. It was found that peer discussions allowed students to articulate their thoughts, enriched their thinking context, and broadened their thinking perspective when collaborating on inferential tasks. Sometimes, the students reached an impasse in a shared activity, so their teacher intervened in the learning activity to facilitate student discussions through free and open exchange of ideas in order to come up with a joint decision to solve statistical problems. After class, each group of the students successfully completed a practical project in which they showed cooperative engagement with inferential tasks and established joint understanding of statistical evidence through collaborative interaction.

Keywords Collaborative interaction · Regression Modelling · Scaffolding assistance · Statistical significance · Vocational education

1 Introduction

Arguments in weather forecasts, opinion polls, and so on are inductive. The arguments about the things that are about to happen around us are usually constructed through statistical inference. The topic of informal statistical inference is taught nowadays in preschool education (McPhee & Makar, 2018) as well as primary and

K. W. Li (✉)
Department of Information and Communication Technology, Hong Kong Institute of Vocational Education (Tsing Yi), Hong Kong, China

M. Goos
School of Education and Tertiary Access, University of Sunshine Coast, Queensland, Australia
e-mail: mgoos@usc.edu.au

secondary education (Doerr et al., 2017), while the topic of formal statistical inference is taught in vocational or undergraduate education. Both types of statistical inference aim to construct reasoned arguments on the basis of statistical evidence, either informally by reading data and finding data relationships, or through formal statistical inference that adopts a more scientific methodology as well as sophisticated statistical tools. Tobias-Lara and Gomez-Blancarte (2019) gave an account of the differences and similarities between these two types of statistical inference.

The topic of formal statistical inference in the present study was taught by the first author from elementary to advanced levels throughout the three years of the Higher Diploma in Applied Statistics and Computing (HDASC) course offered by a vocational education institution. The course aimed at equipping students with statistical knowledge and practical skills for statistical employment. In the first year of study, they learnt how to conduct statistical hypothesis testing by taking the steps of formulating statistical hypotheses; using proper statistical tools for conducting the testing; justifying whether or not to reject a null hypothesis, in turn deducing statistical evidence; and ultimately leading to a conclusion. Statistics modules like Regression Modelling taught in the second year focused on the integration of statistical inference and statistical methods. They were taught how to assess the fit of a regression model ($y = \beta_o + \beta_1 x_1$); how to evaluate statistical significance of the regression intercept (β_o) as well as slope (β_1); and how to make use of the assessment results to generate statistical evidence in association with a claim about the model being feasible or infeasible for making predictions, thus needing to examine the role, strength, and representation of statistical evidence in Regression Modelling (Blume, 2017). In the final year of the study, statistical theories for underpinning the knowledge of statistical inference were taught in relation to how to develop inferential tools; how to detect or compute Types I and II errors arising from hypothesis testing; as well as how to make a trade-off between these two types of errors (see the assignment in Appendix).

Common flaws in students' inferential work were found, such as the wrong formulation of statistical hypotheses, incorrect inferential tools, unsound justification of statistical evidence as well as a mismatch between statistical significance and an argument or a conclusion (Li & Goos, 2013) because statistical inference is tedious and an intellectual activity involving much thinking as well as logical reasoning. For instance, some students were unaware that statistical evidence bearing on a conclusion was not definite because they did not understand the conclusion can only be based on a probabilistic view. That is, there is a chance they drew a wrong conclusion because no statistical tests could achieve 100% significant level. To avoid or rectify the flaws, students should be encouraged to articulate their thoughts, widening their thinking perspective through discussions or debates. This position is supported by Mercer (2004), Vygotsky (1978), and Zavershneva and van der Veer (2018), who argued that a language is a tool for communication and thinking. Students make their ideas available via communication to peers or their teacher. Language can be used for making thinking explicit so that peers and a teacher can read and respond. Students' minds would thus be broadened by appreciating others' verbal responses

and internalizing these as inner speech for directing or redirecting task progression or improvement. Hence, the classroom would be better organized to foster an environment facilitating discussion and social interaction.

2 Literature Review

Research studies, such as those conducted by Goos (2009) as well as Li and Goos (2017) provide us with some insights into sociocultural theories of mathematics and statistics learning. Goos argued that a community of inquiry was established in a mathematics classroom within an IT environment in which students communicated their own beliefs, ideas, and understanding, thus making different contributions and generating a more comprehensive view of learning contexts while the teacher played a role in facilitating learning and provided scaffolding assistance to students. Evidently, the cognitive contribution stemmed from peer collaboration and social interaction among students as well as between students and teacher. On the other hand, Li and Goos reported that the importance of learning partners, social interaction, collaborative learning, the significance of teacher's intervention, and teacher's scaffolding assistance were potential factors influencing social processes of statistics learning within an IT environment. Both studies are relevant to the theme of the present study but the latter report draws closer relevancy to the social context of statistics learning. Nevertheless, it does not give a detailed account of the organization of collaborative learning and the processes of promoting effective peer interaction in a statistics classroom, thus demanding an observational study to address the research question of what patterns of student–student and student–teacher talk are associated with articulation of thoughts about statistical evidence within an IT environment.

Students generally talk while attempting learning tasks collaboratively with their teacher and/or peers. To study the features, nature, and functions of classroom talk, concepts developed by Kumpulainen (1994), Mercer (1995), and Tharp and Gallimore (1995) can be adopted as an analysis framework. Talk among students would be classified by Mercer (1995) as exploratory, cumulative, or disputational. Exploratory talk is developed when students critically evaluate what they are told prior to accepting. Cumulative talk is characterized by students' positive responses without judging by what they are told. Disputational talk features as a peer's proposal being challenged based only on one's personal point of view.

Using Kumpulainen's framework, student talk can be further categorized into (a) intentional—giving gesture, action, voice or sound to express a wish to participate in discussion; (b) responsive—showing one's engagement with learning activities or expressing one's agreement to a lesser extent; (c) affectional—expressing one's personal feelings; (d) reproductional—repeating a peer's response or one's own response without elaboration; (e) experiential—sharing personal experience; (f) interrogative—seeking a peer's feedback when puzzling about their own work; (g) informative—offering facts or knowledge; (h) organizational—doing a tidy-up of ideas, wording or data; (i) compositional—deducing practical implications

for results; (j) imaginative—suggesting rough ideas without any grounds; (k) external thinking—articulating one's thought when talking aloud; (l) judgemental—conveying one's agreement or disagreement; (m) argumentational—challenging a peer's proposal or defending one's argument with evidence; (n) expositional—discovering things that are unfamiliar or unanticipated without detailed planning; (o) heuristic—formulating or regulating strategies; and (p) hypothetical—proposing ideas without elaboration. The first six functions, (a)–(f) are social in nature for initiating and maintaining active communication; the middle four functions, (g)–(j) are mainly for attempting low-collaborative tasks; and the last six functions, (k)–(p) are to promote higher-order thinking through social processes of collective argumentation as described by Brown (2005) as well by Panselinas and Komis (2009).

On the other hand, the talk between students and teacher serves: to elicit knowledge from students; to respond to what students said; and to recap significant ideas or important keywords mentioned by students (Mercer, 1995). Teacher–student talk also serves the purpose of assisting students' performance via these discourse categories: modelling, questioning, cognitive structuring, contingency management, feedback, and instruction (Tharp & Gallimore, 1995). Within each of these discourse categories, the teacher may elicit, respond, or recap. For instance, questioning can be used to elicit what students already know or how their understanding develops or misunderstanding arises. Alternatively, questioning is used for responding to students if the teacher wishes to extend discussions. Questioning may also be used to offer directions towards task improvement or accomplishment after recapping students' useful ideas as hints.

3   Teaching and Learning

To study how peer discussion plays a significant role in statistics teaching and learning within the framework of sociocultural theories of learning, the HDASC Year 2 students (aged 19–22) taking the module on Regression Modelling were divided into small groups in order to increase peer learning and social interaction. They were encouraged to work collaboratively in the classroom and beyond, except in a midterm test and a final examination. They naturally sat together or near each other when attending a class held in a lecture theatre; they collaborated on the worksheets in a practice session conducted in a statistical computing laboratory each week for thirteen weeks. Both the lecture theatre and the laboratory were equipped with computer hardware as well as software, a data projector, microphones, and loudspeakers. All PowerPoint handouts and Excel work were displayed on a projection screen synchronizing with the teacher's talk and verbal and non-verbal cues, for example, using a mouse pointer to draw his students' attention to key areas in Excel programming syntax and results. Instead of delivering lectures, their teacher led class discussions towards developing an understanding of regression topics with practical application using Excel tools.

After the teacher briefly chatted with students in a welcoming tone at the beginning of each lecture, the students went back to the previous handouts while the teacher was reiterating key concepts from earlier topics. This was because the twelve topics in the module were interconnected in the Regression Modelling workflow. In the lecture associated with the present context, students were taught how to conduct statistical hypothesis testing about the significance of model fitting as well as the regression parameters, β_o and β_1 using Excel; how to establish statistical evidence using the test results; as well as how to construct and report reasoned arguments on the basis of the statistical evidence. The teacher initiated discussions inviting all the students to respond. They talked about how they addressed a question of common concern based on their own understandings, opinions, judgements, or perspectives. The discussion they held was similar to collective argumentation or a form of interaction where they joined together to evaluate ideas their classmates brought forward in order to come up with a joint decision leading to a correct and complete answer.

At the beginning of a practice session, the teacher then recapped the concepts of statistical inference and briefly discussed the source, context, and contents of data, and relevant statistical tools on the worksheet he assigned to the students. Each group of the students attempted the learning tasks which were designed to promote an exchange of views, sharing of knowledge, and resolution of problems in order to cultivate a higher level of involvement within a group.

To prepare the students for their prospective statistical careers, the teacher asked students to complete a project on a group basis by the end of the module so as to achieve three educational objectives. First, the project provided opportunities to foster teamwork and cooperative working skills which are essential for the statistics workplace. Second, there is also a need for a project in which students can appreciate the relevance and practical use of Regression Modelling in which they interconnect between statistical concepts; make connections between inferential tools and statistical evidence; construct a Regression Model using the evidence; link the model and the practical context; organize and integrate all of these into a comprehensive report addressing whom it may concern, but, if infeasible, they should use the evidence to substantiate the allegations against model construction. Third, project work was also grounded on sociocultural theories of learning as promoting social interaction among students and enabling verbal exchanges between students to clarify misunderstandings, accomplish tasks, and solve regression problems outside the classroom. Each group could choose one of these themes of study: traffic and public transport, manpower resources, weather, water consumption, retail business, and import and export trades, together with a set of official and relevant data consisting of a dependent variable (y) and three independent variables (x_1, x_2, x_3). The projects were assessed according to the extent to which the students had: (i) scrutinized data; (ii) formulated hypotheses; (iii) utilized inferential tools by means of statistical methods; (iv) evaluated the inferential force of evidence; and (v) constructed reasoned arguments. The first assessment criterion about valuing appropriate and reliable data which would set out the lay of the land of evidence in an inferential process is important (Schum, 2001), while the last four assessment criteria are consistent with the recommendations of Tobias-Lara and Gomez-Blancarte (2019). It is also of great importance in

assessing how the students present logical lines of reasoning in verbal and written modes when making links among concepts, inferential results, and evidence because both modes are essential for communication in the statistics workplace and to display one's thought process.

The projects after assessment were returned to students together with written feedback on their strengths, weaknesses, merits, and/or demerits. The weaknesses or demerits were valuable information about which areas the students should give more attention for improvement and the teacher ought to focus on for instructional scaffolding. In the project reports, many students stated the objectives and fulfilled assessment criteria (i)–(v). Specifically, they constructed credible lines of reasoning about the model building and reported the practical significance and usage of the model together with interpretations of the significant regression intercept (β_o) and slope (β_1).

The students were offered an opportunity to defend and/or clarify the project work in an oral presentation within thirty minutes in the practice session, as enabling each individual student to illustrate his or her own thought process. They went through all the stages in a Regression Modelling process: formulation of project objectives, examination of data, construction as well as validation of a model, and practical implications of the model, mostly together with sound justification and/or statistical evidence. The oral presentations were consistent with their written reports which were mostly elucidative. Their classmates were attentive to the presentation but nobody challenged their work. Only the teacher initiated discussions to promote thinking and reasoning. Statistical evidence was the common thing most students skipped or missed in the point of the argument they wanted to make, and so the teacher asked them to substantiate the argument after eliciting a key phrase in the argument or relevant results to hint at assembling both for the cohesive structure in evidence. If students still could not give a complete answer, the teacher probed for the rationale behind the conclusions after recapping and consolidating their keywords.

4   Observation Study

A study was conducted for observing how peer collaboration might be beneficial to student learning associated with the establishment of statistical evidence. Peer conversation was audiotaped when students were keying in data; programming Excel; reading the screen displays of computer output; and presenting their work on their laboratory worksheets. The students' conversations were transcribed in full, with relevant excerpts being selected for analysis.

In the practice session, the students were given official statistics with a set of social welfare data consisting of a dependent variable (y) and two independent variables (x_1, x_2). They worked collaboratively on computers to attempt four inferential tasks in Regression Modelling. The first task was to fit two regression models to the given data, namely $y = \beta_o + \beta_1 x_1$ and $y = \beta_o + \beta_2 x_2$ using Excel tools. The second and third tasks demanded an evaluation of the significance of the regression intercept

(β_o) and slope (β_1), respectively, in each of the two models using the inferential tool, statistical hypothesis testing. The evaluation task is about examining the strength of statistical evidence. The fourth task following the previous two tasks was to report a mathematical relationship between a dependent variable and an independent variable for each of the two models that would make sense after β_o and/or (β_1) had achieved statistical significance.

Interactions between three students (with codes, J, K, and L to remain anonymous) are analysed to illustrate the effects of peer collaboration. L and K had already formed a group, and J joined because her fellow group member was absent from the practice session. The conversation of only this group of three students was available for analysis for two reasons. First, few students agreed to voluntarily participate in the observation study. Second, the microphones for audiotaping were placed at a distance from those who did agree to participate so as not to disturb them from doing their work; however, some still tended to talk less frequently and less audibly because of being anxious about having their conversations audiotaped. Thus, despite following all ethical processes for conducting the research, only one group's audio-recording was available for analysis.

Prior to attempting the tasks, students judged whether the data were relevant and credible as well as whether the measurement and measurement units of the data covered a reasonable and meaningful range. This judgement takes precedence over any statistical methods to avoid discrediting inferential results (Schum, 2001). They accomplished the first task efficiently because they had become proficient at building Regression Models after attending lectures and accomplishing similar tasks in the previous few practice sessions. Shortly after, they read the Excel output in which they checked regression estimates of β_o and β_1, together with significance testing results so as to attempt the second and third tasks. They had trouble with grading the strength of statistical evidence because they could not distinguish between two available approaches to evaluating statistical evidence. The first approach is to compare the test statistic with the critical value at a pre-specified level of significance in order to decide whether or not to reject H_o (a null hypothesis). Another approach deals with checking whether or not the decision substantiates or forms persuasive lines of reasoning based on the p-value which reports the chance of committing a Type I error; i.e., the p-value needs not fix a certain level of significance (α) but evaluates how likely it is that a true hypothesis would be rejected. Both approaches are equivalent to a certain extent in testing of statistical hypotheses. The first approach is relatively simple and straightforward but the second approach is relatively more objective. Kumpulainen's (1994) and Mercer's (1995) frameworks were used to analyse the nature and contents of talk among students (Tables 1–4) when they moved on to selecting and making use of an appropriate statistical tool.

Both students, L and K were oblivious of what output was produced by the Regression Analysis tool in Excel. Their queries were concerned with the proper selection of a statistical tool, the test statistic versus p-value. L's talk was interrogative, characterized by seeking her peers' approval for the use of the p-value because L was concerned about instances in which they attempted to evaluate the extent to which H_o would be rejected or not. J's response simply agreed with L's proposal, as believing

Table 1 The nature and contents of talk among students, J, K, and L when selecting an appropriate statistical tool

Excerpt	Student code	Contents of talk
1	L:	Do we just look at the p-value? (**Interrogative**)
2	J:	Yes (**Responsive**)
3	L:	We are going to use Excel (**Responsive**)
4	K:	Are we allowed to use p-value? (**Interrogative**) Does the teacher ask us to compute? (**Interrogative**)
5	J:	The teacher doesn't want us to compute but directly make use of p-value (**Responsive**)

the p-value is a proper tool, then L's subsequent responsive talk sidetracked the issue and proposed to use Excel. Meanwhile, K was puzzled, so K's interrogative query brought her peers back to the question of using the p-value and raised another question of whether to compute the test statistics and compare with α. J clarified by recalling the teacher's recommendation and she also suggested making use of the p-value resulting from Excel output without the need for computing. The recommendation about the widely adopted tool, p-value for justifying statistical evidence in hypothesis testing was briefed at the beginning of the practice session. According to Mercer (1995), the above excerpts can be categorized as cumulative because their talk merely exhibited a simple question-and-answer form focusing on the selection of statistical evidence to reject or not to reject H_o, but without involving critical evaluation or tool justification. The students eventually made use of the p-value to grade the extent to which H_o would be rejected (Table 1).

Student K recalled the statistical logic that is critical for evaluating H_o in deciding whether or not a regression parameter, β_o or β_1 was statistically significantly different from zero. She checked the Excel output and found the p-value was smaller than α and immediately announced, *"A smaller p-value indicates to reject H_o. That should be the case"*, implying rejection of H_o. Her utterance (Excerpt 6) displayed external thinking. J's talk seemed argumentational in proposing to use another statistical tool, the t-statistic (i.e. the value resulting from the test statistic), and outlined the rejection criterion for H_o in which she mixed up p-value, α, and t-statistic. L responded by attempting to rectify J's mistakes, but did so neither completely nor exhaustively (Excerpt 8). K insisted on using the p-value as a rejection criterion and recalled rejection and non-rejection criteria of H_o thoroughly before deciding to reject or not to reject H_o (Excerpt 9). Her utterance was therefore informative. There was no more negotiation about the rejection criterion and they eventually agreed to use the p-value (see Table 2).

Using Mercer's (1995) framework to analyse the above excerpts, the talk was probably exploratory although they did not negotiate much about the rejection criterion. In fact, they did not totally accept their peer's proposals but evaluated their own ideas critically and also raised counter-proposals.

Table 2 The nature and contents of talk among students, J, K, and L when checking the criteria for rejecting H_o

Excerpt	Student code	Contents of talk
6	K:	p-value is larger, … A smaller p-value indicates to reject H_o. That should be the case (**External thinking**)
7	J:	If t(-statistic) is larger than α, … p-value? (**Argumentational**)
8	L:	(p-value is) smaller (than α means to reject H_o) (**Responsive**)
9	K:	(If p-value is) smaller (than α, that) is to reject (H_o), larger (than α) is to accept (H_o); (or say) not to reject (H_o). It is very confusing (**Informative**)

While checking the p-value resulting from a statistical test, J did not understand the p-value being presented using scientific notation and asked, *"E?" I don't know what E is?"*, so her talk was an interrogative enquiry about the mathematical notation. L replied that *E* represented exponential notation, using external thinking. They all subsequently could read the p-value resulting from the statistical test and their talk was expositional when announcing a small p-value (Excerpts 12, 13, and 16). A small p-value here concerns the level of statistical significance for which we would reject H_o. Specifically, J pointed out that the p-value is 10 to the power negative twelve (i.e., … $\times 10^{-12}$) (Excerpt 17). J and K decided to reject H_o, thus forming a basis for a valid statement, a probable conclusion, or an inductive argument (Excerpts 14 and 15). All these three excerpts represent external thinking involving articulation of thought.

This part of peer talk (Table 3) is related to reading a statistical result from a screen display of Excel outputs as being visible products of problem-solving from students' joint effort (Mercer 2005). Statistical results output from Excel initiated processes of thought and action in a physical form shared in their collaborative work. Their utterances are classified as exploratory according to Mercer (1995) because

Table 3 The nature and contents of talk among students, J, K, and L when grading the strength of statistical evidence

Excerpt	Student code	Contents of talk
10	J:	E ? I don't know what E (in the p-value) is? (**Interrogative**)
11	L:	E represents exponential (**External thinking**)
12	K:	This is very small (**Expositional**)
13	J:	This is very small (**Expositional**)
14	K:	That is to reject (H_o) (**External thinking**)
15	J:	Reject (H_o) (**External thinking**)
16	L:	(4E-12 is) so small! (**Expositional**)
17	J:	This E(-12) is 10 to power negative twelve (**External thinking**)

Table 4 Functions of talk displayed by the three students

Function of talk	Frequency		
	Student J	Student K	Student L
Interrogative	1	1	1
Responsive	2		2
External thinking	2	2	1
Argumentational	1		
Informative		1	
Expositional	1	1	1
Overall	7	5	5

they all critically evaluated the p-value in connection with H_o individually and J and K concluded to reject H_o.

All the students were engaged with peer discussion (Tables 1, 2 and 3) in which the question and answer exchanges shaped the flow of inferential tasks on which they worked. Table 4 summarizes various functions of talk displayed by them in order to generate a more comprehensive view of thinking when attempting the learning tasks together. The patterns of talk displayed by the three students, J, K, and L were similar; much of their talk was external thinking, argumentational, and expositional in nature for examining the strength of statistical evidence using higher-order thinking. Some of their talk had the purpose of expressing agreement to a lesser extent when selecting the tool for checking statistical evidence, and only one instance was for attempting a low-collaborative task as in summarizing the criteria of H_o rejection.

When the students were struggling to understand the scientific notation of the p-value, the teacher intervened and elicited J's response, "4×10^{-12}". He then modelled for them how to read an extremely small p-value in scientific notation that is equivalent to a very low probability of H_o being true. To progress in model building requires a decision criterion of whether or not to reject H_o at a given level of statistical significance, α. The teacher also modelled the selection of the level of significance, 5% that is commonly adopted as an acceptable level of Type I error in statistical inference because it entails a rational weighing of the evidence; otherwise, using a higher level of statistical significance would diminish the credibility of the evidence. The teacher further modelled the physical act of searching for the p-value that would be used as statistical evidence when contrasting with α. J's response in Excerpt 27 affirmed the rejection of H_o. The dialogue between the students and teacher displayed verbal exchanges. The teacher elicited their written as well as verbal responses, which they supplemented after the teacher had recapped their incomplete responses. Modelling assistance was offered successfully to associate with the level of statistical significance and extend their discussion (refer to Table 5).

A decision in favour or not in favour of rejection of H_o yielded opposite implications for the significance of a regression parameter, β_o or β_1 as well as the regression model being built. For this reason, the teacher checked how the students formulated H_o in association with the role and representation of statistical evidence. Student

Table 5 The contents of talk between the teacher and students, J, K, and L when reporting and grading the strength of statistical evidence

Excerpt	Teacher and Student code	Contents of talk
18	Teacher:	Yes, you can just write it down, 0.0000 because you have a value 4×10^{-12} which is zero point … and then we have …
19	J:	Twelve zeros
20	Teacher:	Twelve zeros … and then we have 4. So, if we look at four decimal places which are similar to 0.0000 (after truncation). Understand what I mean? Because we compare with 0.0 …
21	J:	5
22	Teacher:	5 so obviously this (p-) value is much …
23	J & K:	Smaller
24	Teacher:	Smaller? or larger?
25	J & K:	Smaller
26	Teacher:	(4×10^{-12} is) much smaller than 0.05, so we reject Ho
27	J:	Yes!

J gave a quick but incorrect response due to misinterpreting the teacher's question rather than making any conceptual error. Her instantaneous response, "*Sorry! Sorry!*" indicated immediate self-awareness of her mistakes as not being approved by her verbal thought. While she was attempting to correct her H_o formulation, L interjected, "*H_o: $\beta_o = 0$*" and her prompt H_o reformulation was much quicker than the teacher's correction. The teacher recapped L's correct answer to reinforce the concept of H_o. The three students, J, K, and L were laughing and their laugh was sincere as their stress at getting the correct H_o formulation was being released. The teacher attempted to extend discussion of the formulation of H_1 (alternative hypothesis), without success (Excerpt 32). However, his modelling of the formulation of H_1 allowed J to respond to the cue, "*H_1: $\beta_o \neq 0$*". He then gave feedback to validate her correct answer (Table 6).

The teacher initially offered the three students, J, K, and L cognitive structuring assistance by providing explanatory structures for justifying the rejection decision based on the p-value. They were confused about deducing the implication of rejecting H_o so the teacher asked, "*What does it mean? … We reject H_o by looking at two different values (p-value and α-value)?*" They found the question vague, so the teacher rephrased the question more as a statement, "*Does it mean this small (p-) value is in the rejection region, isn't it?*" J replied, "*Yes!*" To reinforce the concept of rejecting H_o the students already held, the teacher modelled drawing a conclusion from the statistical evidence when contrasting between the areas corresponding to the p-value and α. The teacher found they could gradually pick up the responsibility for the tasks so he then left them alone to draw a conclusion that was their resulting belief in the falsehood of H_o. The conclusion based upon probabilistic

Table 6 The contents of talk between the teacher and students, J, K, and L when formulating H_o and H_1

Excerpt	Teacher and student code	Contents of talk
28	Teacher:	What is H_o in this case?
29	J:	H_o is 0.05. Sorry! Sorry! *(While J was saying "Sorry!" L shouted, H_o: $\beta_o = 0$. Both students admired L's answer and said, "Very good!")*
30	Teacher:	H_o: $\beta_o = 0$ *(The three students were laughing.)* H_1 ... What is H_1?
31	J:	H_1 is …
32	Teacher:	H_1: $\beta_o \neq …$
33	J:	H_1: $\beta_o \neq 0$
34	Teacher:	Not equal 0, yes! OK?

belief is not completely credible (Schum, 2001). The students were actively engaged in discussion in which the teacher played key roles of orchestrating social interaction between students; facilitating student discussion; and offering direction towards deeper thinking (Table 7).

Table 7 The contents of talk between the teacher and students, J, K, and L when discussing two approaches to evaluating statistical evidence

Excerpt	Teacher and student code	Contents of talk
35	Teacher:	So, if a very small (p-)value, much smaller than 0.05, we can reject H_o
36	J, K & L:	Hm!
37	Teacher:	What does it mean? We reject H_o by looking at two different values (p-value and α-value)? Does it mean this small (p-)value is in the rejection region, isn't it?
38	J:	Yes!
39	Teacher:	This (p-value) is very small because the rejection region is larger than the (p-)value … *(The teacher pointed at the p-value displayed on their computer monitor.)*
40	J:	The (p-)value
41	Teacher:	We have obtained, OK?
42	J:	I see! *(The teacher left them alone to draw a conclusion.)*

The students continued to work together without further intervention from the teacher. They successfully completed the last task by reporting a mathematical relationship between a dependent variable and an independent variable for each of the two models they had built.

Apart from collaboratively accomplishing inferential tasks in the statistical computing laboratory, students L and K also submitted their project report for assessment; student J did not do so because J was in another project team. In the report, both L and K spelt out the objectives of their project; made judicious use of the data available to them by means of descriptive statistics and statistical graphing as convincing arguments are built on the relevant, reliable, and valid data (Schum, 2001) free of bias (Shield, 2000). They utilized graphical and computational tools to establish sound evidence in model building. They had an incorrect suspicion of outlying observations, which in fact did not exist, so neither regression problems were created nor the model fitting was distorted. In model validation, they verified the linearity, homoscedasticity, independence, and normality of the data. They elaborated the practical usage of the model and the meaning of the regression estimates of β_o and β_1 relating to the context and measurement units of data. A conclusion on the practical significance of the model for making predictions was made. More importantly, the lexical content in the report showed logical flow, thinking, and reasoning.

In the project presentation, student L gave an introduction; clearly stated the project objectives; confirmed the source of data (i.e., y, x_1, x_2, x_3) was reliable and the data content was relevant; checked the measurements of the data were meaningful and unbiased by utilizing the measures of central tendency as tools. She proposed to build three Regression Models for making predictions of y from x_1, x_2, and x_3 on the grounds that scatterplots exhibited a linear relationship between y and x_1; y and x_2; and also y and x_3 and she further substantiated the relationship using the correlation coefficient. Among the three models, she suggested using the model with the largest correlation coefficient. Her peer, K took turns in justifying and comparing the model fitting among the three models using R^2 (the coefficient of determination indicating the goodness of fit of a model) and decided to choose the model with the largest R^2 and further illustrated the best model fitting using a graph of y against both observed and predicted values of x. The model was affirmed by verifying the regression assumptions and non-existence of outlying or influential observations. Furthermore, she showed that the Regression Model as well as the intercept (β_o) and slope (β_1) of the model were significant together with statistical evidence; gave a practical interpretation of both estimates of β_o and β_1; and ultimately concluded that the model was feasible for making predictions in connection with the data context. Both students gave a clear verbal presentation of the modelling workflow without any interruption and utilized statistical evidence to defend the model they had built. Each of them played a significant role and gave contributions to the project work.

Nevertheless, the teacher posed questions to probe the students' construction of logical lines of reasoning on statistical evidence. L was attentive to the first question but gave ambiguous responses, so the question relating to the notion of Regression Modelling of the relationship between y and x was raised by the teacher after giving a cued elicitation, i.e. the correlation coefficient presented by the students. L could not

directly answer the question but highlighted correlation concepts. After the teacher had rephrased his question directing her towards the data context of *y* and *x*, she managed to give a complete answer. Another question was raised to justify the model fitting based on the graph, and K attempted to answer but was incomplete. L gave her some help and asked the teacher to elaborate on the question. She gave responses without directly addressing the question. Teacher questioning continued and two hints were given. One was about decoding observed and predicted values of *x* on the graph and the other was about drawing comparisons between these two values. She followed the hints to construct an explicit argument in defence of the model fitting. Both students worked in close collaboration on the project, shared the workload, and offered mutual assistance. The content of verbal exchanges between the students and teacher has a rich thinking context.

5   Conclusion

Evidence deriving from statistical results is by no means perfect, owing to some degree of uncertainty; the credibility of statements, arguments, or conclusions based on the evidence may be challenged from various perspectives. To ascertain credibility, it would be necessary to avoid personal bias by having input from more people embracing a broader view of thinking to form intellectual resources for sharing such that lines of reasoning can be constructed or revised exhaustively. This present study shows the cohesive structure of discussion that develops conceptual understanding of statistical evidence, mostly by examining the role, strength, and representation of statistical evidence (Blume, 2017). Specifically, cumulative talk was delineated by the pattern of simple responses given to interrogative enquiry about tool selection. Exploratory talk was observed when involving articulation of thinking and discovering the extraordinarily small p-value using external thinking, followed by expositional speech. Both patterns were evident from verbal exchanges among the students; with the second pattern displaying higher-order thinking. Exploratory talk was necessary for them to critically evaluate the p-value in connection with H_o using verbal expositions: external thinking, along with argumentational, responsive, and informative talk. Irrespective of the types of talk they used, the group of students gained from collaboration and subsequently generated a more comprehensive view of thinking so as to complete a statistical inference.

However, statistical inference required thinking and reasoning that sometimes could not be developed only through peer discussion and joint performance of tasks in the absence of the teacher. The teacher therefore regularly intervened to check students' learning progress, promote intellectual exchanges between them, and offer them learning assistance via modelling, cognitive structuring, and feedback. Modelling assistance was offered to stimulate students' thinking and provided direction in advancing the level of their thinking in association with the tasks of formulating the null and alternative hypotheses; selecting the level of statistical significance; constructing statistical evidence; and drawing conclusions from the

evidence. The teacher elicited knowledge from the students and recapped their incomplete verbal responses, but also offered useful and important modelling assistance. Cognitive structuring was adopted to organize students' thinking associated with the inferential tasks of justifying the rejection decision based on the p-value. The teacher gave feedback to validate students' correct responses. A similar approach to assisting students in achieving specific learning objectives was evident from students' oral project presentations.

The findings of these analyses of talk in a statistics classroom and beyond are grounded in sociocultural theories of learning and revealed how peer discussion structured thought in the context of statistical evidence. However, conclusions are only tentative since not all students in the class volunteered to participate in the observational component of the study (whereas all agreed to participate in the questionnaire-based survey and experimental study components). Among the student volunteers, some appeared to be anxious at the prospect of having their conversations audio-recorded, and as a result, talked less frequently and less audibly. Hence, the conversation of only one group of three students was available for analysis; however, based on the teacher's knowledge of this class, it is assured that they are representative of the kind of talk observed throughout the observation period.

Appendix

The following is a problem scenario in the Statistical Inference assignment in HDASC Year 3.

A random sample of 30 shoppers will be chosen in a supermarket to try a new flavour of ice-cream. They will be asked whether they like it or not. The marketing department of a dairy products company wants to know whether the new flavour of ice-cream will have a successful launch based on this market testing. Let p denote the proportion of shoppers who like the new flavour and let the test statistic X represent the number of shoppers who like the new flavour. Suppose that you are a statistical officer assigned to this project.

Complete the following tasks:

(i) For a successful launch, at least half of the shoppers should like the new flavour. State H_0 and H_1 about p.
(ii) Determine the value of the Type I error when $X \leq 15$ is chosen as the rejection region.
(iii) Determine the value of the Type II error when $p = 0.4$ represents the true proportion of shoppers who like the new flavour.
(iv) Use Microsoft Excel to complete the following;

    (a) design and implement a spreadsheet that can calculate the values of the Type I and Type II errors for different rejection regions and

    (b)   construct **ONE** chart for a comparison of the values of the Type I and Type II errors resulting from the spreadsheet you gave as your answer to (iv)(a).

(xxii)   Based on the chart you gave as your answer to (iv)(b), describe briefly the relationships between:

    (a)   the Type I error and the test statistics and
    (b)   the Type II error and the test statistics.

(vi)   Based on the chart you gave as your answer to (iv)(b), determine the best rejection region to be used for hypothesis testing for the company, giving reasons for your answer.

References

Blume, J.D. (2017). *Examining the role and representation of statistical evidence in science.* Retrieved from, https://www.statisticalevidence.com/.

Brown, R. (2005). Exploring spaces of collaborative learning within a collective argumentation classroom. In M. Goos, C. Kanes & R. Brown (Eds.), *Proceedings of the 4th International Mathematics Education and Society Conference* (pp.104–114), Gold Coast, Australia: Mathematics Education and Society.

Doerr, H. M., Delmas, R., & Makar, K. (2017). A modelling approach to the development of students' informal inferential reasoning. *Statistical Education Research Journal, 16*(2), 86–115.

Goos, M. (2009). A sociocultural framework for understanding technology integration in secondary school mathematics. In M. Tzekaki, H. Kaldrimidou & H. Sakonidis (Eds.), *Proceedings of the 33rd Conference of the International Group for the Psychology of Mathematics Education* (pp.113–120), Thessaloniki, Greece: Psychology of Mathematics Education.

Kumpulainen, K. (1994). Children's talk during collaborative writing at the computer. *Reading, 28*(2), 6–9.

Li, K.W., & Goos, M. (2013). Evaluating students' correlation graphing capability using SOLO taxonomy. In: *Proceedings of the 59th ISI World Statistics Congress* (pp. 3411–3416). The Netherlands: The International Statistical Institute. Retrieved from, https://2013.isiproceedings.org/Files/CPS008-P5-S.pdf.

Li, K. W., & Goos, M. (2017). Factors influencing social processes of statistics learning within an IT environment. *International Journal of Science, Mathematics and Technology Learning, 24*(2), 21–33.

McPhee, D. & Makar, K. (2018). Early childhood experiences in inferential statistics using an inquiry approach. In M. A. Sorto, A. White & L. Guyot (Eds.), *Proceedings of the 10th International Conference on Teaching Statistics*. The Netherlands: The International Statistical Institute. Retrieved from, http://icots.info/10/proceedings/pdfs/ICOTS10_4B1.pdf.

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

Mercer, N. (2004). Sociocultural discourse analysis: Analyzing classroom talk as a social mode thinking. *Journal of Applied Linguistics, 1*(2), 137–168.

Panselinas, G., & Komis, V. (2009). Scaffolding through talk in groupwork learning. *Thinking Skills and Creativity, 4*(2), 86–103.

Schield M. (2000). Statistical literacy and mathematical thinking. In: *Paper Presented at the 9th International Congress On Mathematics Education*. Tokyo, Japan. Retrieved from, http://www.statlit.org/PDF/2000SchieldICME.pdf].

Schum, D. A. (2001). *Evidential foundations of probabilistic reasoning*. Northwestern University Press.

Tharp, R. G., & Gallimore, R. (1995). *Rousing minds to life: Teaching, learning, and schooling in social context* (Reprinted). Cambridge University Press.

Tobias-Lara, M. G., & Gomez-Blancarte, A. L. (2019). Assessment of informal and formal inferential reasoning: A critical review. *Statistical Education Research Journal, 18*(1), 8–25.

Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.

Zavershneva, E., & van der Veer, R. (2018). *Vygotsky's notebooks*. Springer.

Dr. Ken W. Li is Retired Lecturer and serves on various editorial boards of international academic journals as well as conference proceedings, such as Associate Editor of Electronic Journal of e-Learning, UK. He was General Secretary of the Hong Kong Statistical Society. His research interests include computer simulation, big data analytics, and the use of technology in education. He has published research papers in the areas of computer science, statistics as well as education and co-authored an undergraduate textbook.

Prof. Merrilyn Goos is Professor of Education at the University of the Sunshine Coast, Australia. She is a mathematics educator whose research has investigated students' mathematical thinking, the impact of digital technologies on mathematics teaching and learning, and the professional learning of mathematics teachers and mathematics teacher educators.