


# An Experimental Study on Peer Collaboration and Student Performance in an IT Environment

Ken W Li <sup>1\*</sup> , Merrilyn Goos <sup>2</sup> 

<sup>1</sup>Hong Kong Institute of Vocational Education, CHINA

<sup>2</sup>The University of Queensland, AUSTRALIA

\*Corresponding Author: [dr\\_ken\\_li@yahoo.com.hk](mailto:dr_ken_li@yahoo.com.hk)

**Citation:** Li, K. W., & Goos, M. (2021). An Experimental Study on Peer Collaboration and Student Performance in an IT Environment. *International Electronic Journal of Mathematics Education*, 16(2), em0632. <https://doi.org/10.29333/iejme/10890>

## ARTICLE INFO

Received: 10 Jul. 2020

Accepted: 9 Jan. 2021

## ABSTRACT

This paper addresses the question of whether peer collaboration affects students' performance of regression modelling tasks, an experimental study consisting of a test was conducted in a computing laboratory. Collaborating groups of students were randomly assigned to one of three experimental conditions: pre-task discussion (i.e., group members had discussions right at the beginning of the test), post-task (i.e., group members had discussions in the middle of the test) or no discussion. It was found that the influence of the pre-task discussion on a preliminary examination of regression data was not very explicit in accomplishing tasks involving judgement of data reasonableness and meaningfulness, correlation calculation, Excel programming, and statistical hypothesis testing. Yet, students who had pre-task discussion showed better quality in graph construction as well as graph characterization and also performed better in correlation interpretation and reasoning tasks than those who did not. Students could clarify their misunderstanding and/or refine their thoughts through post-task discussion but no improvements in comprehension of data were detected for correlation calculation; Excel programming, hypothesis testing, and correlation deduction as well as synthesis.

**Keywords:** correlation, pre-task discussion, post-task discussion, SOLO scores, statistical graphing

## INTRODUCTION

Statistics is not well grasped by many students for various reasons. Dunn et al. (2016) reported that the language of statistics is complex and difficult to teach and learn owing to discrepancies in the meanings of words when used in different contexts, such as statistical contexts versus English in daily usage, or contexts other than statistics, e.g., engineering, medicine, psychology, and business. In addition, there are no universal statistical terms or notations common to all statistical software and textbooks. Besides, many students hold negative attitudes towards statistics learning, thus raising learning barriers and eventually adversely affecting their statistics achievement (Bude, 2007; Schau, Millar, & Petocz, 2012). The study of teaching and learning of statistics is thus of paramount importance and has a wide coverage. Efforts to update and improve statistics curricula are increasing in schools and higher education. From mathematics educators' perspective, Watson and Neal (2012) looked into the new Australian secondary school curriculum that would fulfil the growing needs for statistically literate people who are able to locate, analyze, and use information in their daily lives and/or their employment. From a statistics education perspective, Horton and Hardin (2015) updated the undergraduate statistics curriculum to prepare students for their statistical careers so they can meet the challenges of an information age. Research has also been devoted to students' statistical and probabilistic thinking, statistical reasoning, and communication skills (Dunn et al., 2016; Gordon & Finch, 2010; Malaspina & Malaspina, 2020; Pfannkuch, 2011; Pfannkuch & Wild, 2004). Tremendous efforts were put into enhancing statistics teaching and learning through various pedagogical approaches and technology (e.g., Pfannkuch & Budgett, 2016; Wild et al., 2015).

The education of pre-service teachers and professional development of in-service mathematics teachers are also crucial to student learning but some teachers could not demonstrate a full grasp of some statistical concepts (Biehler, 2016; Casey & Wassermann, 2015; da Silva, Kataoka, & Cazorla, 2014; Reaburn, 2017). Some studies (e.g., Biehler, 2016; Fernandez et al., 2020; Haines, 2015) have already addressed this issue.

Despite these research outputs manifesting a macro view for improving pedagogical content knowledge in statistics education, many studies still report that students find it difficult to learn statistics. For example, many students misconceive correlation as being causation and having transitivity (Sotos et al., 2009). Although there has been recent interest in the influence of culture on students' understanding of statistics (e.g., Sharma, 2014), few studies have closely examined the social processes of statistics learning that might stimulate students' thinking and broaden their perspectives on problems via peer collaboration (see

Curran, Carlson, & Celotta, 2013). This paper thus attempts to address the following research question: *How does peer collaboration in an information technology (IT) environment affect students' performance of statistical thinking and statistical graphing tasks in correlation analysis?* IT here offers capabilities beyond educational delivery of teaching and learning materials or facilitation of statistical analysis and computation, but has much to do with re-organizing an environment in which students among themselves develop learning partnerships to socially construct knowledge through peer collaboration and discussion as well as social interaction.

## THEORETICAL FRAMEWORK

The theoretical framework for the study was derived from socio-cultural theories of learning, statistical literacy, and the pedagogy for teaching correlation with an emphasis on developing statistical thinking and graphing in a collaborative learning environment for exchanges among students when evaluating statistical information critically and appreciating statistical evidence.

### Socio-cultural Theories

According to Vygotsky's socio-cultural theory (1978), learning is a social process in which language is a tool for communication as well as thinking through which students interact among themselves or with their teacher. They make their ideas available via communication to others for comment, suggestion, and argument as well as students' thinking is further developed by having to make sense of what others say to them. Irrespective of whether students substantiate their own claims or challenge their peers' proposals, their thoughts are articulated and ideas and concepts will become more refined.

Diverse applications of socio-cultural theory to educational research provided insights into building social environments in classroom teaching and learning (Goos, 2004, 2009, 2014; Mogro-Wilson, Reeves, & Charter, 2015; Wang, Bruce, & Hughes, 2011). Mogro-Wilson, Reeves, and Charter (2015) investigated a teaching strategy for developing statistical abilities of doctoral social work students by organizing a learning environment that would facilitate knowledge exchanges among students. It was speculated that students working collaboratively may have access to alternative perspectives on problems at hand through reviewing and interpreting their peers' feedback. On the other hand, Wang, Bruce, and Hughes (2011) applied socio-cultural theory in information literacy research for the establishment of a collaborative culture within a community in which university academic and technical staff gave opinions on how to integrate the course contents of Information Literacy into different undergraduate programs from diverse perspectives. In these two studies, socio-cultural theory served different academic purposes, but neither study reports the detailed practices of teaching and learning in the classroom.

In contrast, Goos (2004, 2009, 2014) drew closer to the theme of the present study, using a socio-cultural framework to report a finer-grained account of teaching and learning practices using technology in secondary mathematics classrooms. First, a community of inquiry was established in the way that the teacher clarified misunderstandings, structured students' thinking, and moderated discussion among students. Through discussions, the students came to see the same problem differently and proposed interpretations of problem settings leading to different approaches to problem-solving. To respond to their peers' feedback or different approaches, they communicated their own beliefs, ideas, and understanding, thus making different contributions and generating a more comprehensive view of the problems they were asked to solve (Goos, 2004).

Second, Goos (2009) switched to study the factors influencing teachers' use of technology in secondary mathematics classroom using the theories about ZPD (the zone of proximal development) (Vygotsky, 1978) as well as ZFM (the zone of free movement) and ZPA (the zone of promoted action) (Valsiner, 1997). The ZPD is more about a teacher's personal abilities, mathematical knowledge, and pedagogical beliefs as well as content knowledge. The elements of ZFM include students' abilities, motivation, and behavior; curriculum, teaching materials, and assessment; and school cultures, resources, and support. Pre-service teacher education, professional development, and interaction with teaching colleagues are the elements of ZPA. There is also an interaction effect among these three zones on a teacher's willingness to integrate technology into classroom teaching. Goos (2014) further argues that these three zones would influence the development of a teacher's pedagogical identity in relation to confidence and beliefs about managing class time and resources as well as adapting and responding to the changes in curriculum and assessment requirements when integrating technology in classroom teaching and learning.

Within a socio-cultural perspective, Goos (2004) focused on the learning process within a community of inquiry, whereas both studies (Goos, 2009, 2014) direct attention to teacher's knowledge, ability, and pedagogical identity as well as teacher-environment relationships within school context. The ability refers to a teacher's skill and/or experience in working with technology. The environment is about social and physical conditions or settings that a teacher encounters when interacting with students, technology, resources, and tools.

### Statistical Literacy

Statistical literacy refers to the extent to which a student has demonstrated ability to read and evaluate statistical information used in arguments (Carter & Milne, 2000). Gal (2003, p. 16) defined statistical literacy as "the ability to interpret, critically evaluate, and express one's opinion about statistical information and data-based messages." Both definitions manifest the core elements of statistical literacy: comprehension, evaluation, and judgement. Statistical literacy takes on additional elements in the definitions given by Garfield (2011), Grant (2017), and Watson (2006) by enlisting statistical tools and statistical communication.

Interestingly, the definitions of statistical literacy have been reviewed regularly by Schield (1998, 1999, 2001, 2002, 2010, 2014) to extend its scopes over the past twenty years. Schield's (1998) preliminary definition of statistical literacy targeted descriptive

statistics, statistical inference, Bayesian statistics as well as evidential statistics. To articulate the definition, statistical literacy involves an ability to comprehend as well as interpret statistical information, and to think critically about arguments using statistics as evidence (Schield, 1999). The sources of variation – chance, bias, and confounding – are taken into consideration and must be explicated when making a statistical claim (Schield, 2000). Reading, comparing, and interpreting data presented in tables, charts, and so on correctly are challenging tasks that form part of statistical literacy (Schield, 2001, 2010, 2014). The term “statistical literacy” as used by Schield (2001) refers to equipping students with common statistical tools: the measures of central tendency and dispersion of data, graphical displays and tabular presentation of data, and p-value for handling statistical tasks or matters. Although everyone should be statistically literate, Schield (2002) further discusses chance-based literacy, fallacy-based literacy, and correlation-based literacy. Specifically, correlation-based literacy is essential for the preliminary stage of regression modelling in the present study.

Irrespective of the wordings used in the above definitions of statistical literacy, they have commonalities. Broadly, a person should demonstrate common sense in the process of inquiry, while reasoning critically about using statistics as evidence. Specifically, a person should be able to read, interpret, critically evaluate, and communicate statistical information; properly utilize statistical tools; present statistical findings in meaningful ways; as well as reason about data-related claims or chance-related phenomena. In spite of the usefulness of listing these elements, one could also argue for the need to identify and organize the relationships between them in order to operationalize the definitions more efficiently. In line with the framework developed by Li (2006), and Pierce et al. (2014), a synthesis of the definitions thus proposes that statistical literacy operates at three levels of increasing sophistication – comprehension, planning and execution, and evaluation. Level one requires comprehending data – reading factual information and grasping the implicit meaning of statistics presented or published by some other person. This can go beyond achieving surface understanding of data to evaluating statistical information by finding relationships between data and screening information that is relevant to one’s own statistical tasks; as well as appreciating information arising from probabilistic or stochastic phenomena. Comprehension tasks take precedence over statistical methods organized in Level two that involves planning statistical tasks and their execution. Planning and execution of statistical tasks requires the exercise of thinking, reasoning, statistical graphing, and statistical communication. Level three refers to evaluating the context, the power, and the limitations of reasoned arguments on the basis of statistical evidence; and how well or how poorly statistical results match the real-life context.

### **Statistical Pedagogy**

Although the hierarchy of statistical literacy arranged into the three levels of comprehension, planning and execution, and evaluation is useful for educating students, the variety of statistical knowledge required ranges from simple to complex according to the academic levels of students and the nature and complexity of statistical tasks. To address the question of pedagogy for teaching vocational students the topic of correlation analysis, it is imperative to help students develop statistical thinking and graphing abilities.

Comprehension here requires reasoning about data that takes precedence over statistical methods by looking beyond numerical representation and judging whether or not the measurement and measurement units of data cover a reasonable as well as meaningful range. The students may gain some insight into where data come from through deriving personal meaning.

Planning and execution of correlation tasks requires the exercise of scatterplot construction, and thinking associated with correlation comprehension, statistical calculation, and hypothesis testing. Evaluation refers to justifying whether or not the scatterplots show the linear relationship between two variables, and deducing the relationship between two variables. Prior to evaluation, students must understand the context of the data and the implications of statistical results.

The pedagogy for teaching correlation analysis aimed at improving classroom teaching practice with an emphasis on social processes of learning. A cognitive model of correlation comprehension was developed by Li and Goos (2011), arising from a synthesis of three different perspectives: pedagogy, statistics, and cognitive psychology. The model starts with pattern recognition processes, followed by interpretive processes and then integrative processes to accomplish the remaining tasks. The pattern recognition process commences by checking the data encoded on a scatterplot. Interpretive processes are perceptual processes that operate on those patterns to retrieve or construct qualitative and quantitative meanings. Integrative processes are conceptual processes that relate the meanings to the graphic features, such as titles, labels, and scales or plotting symbols in a graph.

To increase students’ opportunities for peer learning and collaboration, students were divided into collaborating groups of two in computing laboratory sessions. In order to reduce the extent of academic variability among collaborating groups, consideration was given to ability compositions in groups. Thus, a less competent student was generally grouped with a more competent peer. These groups become more homogeneous in terms of students’ academic abilities on the basis of their grade point averages achieved in their Year 1 Study, and members within each of these groups were expected to work together in and after their classes. This would enable a more competent student to assist his or her less competent learning partner, thus creating the necessary conditions for observing whether or not peer assistance might be beneficial for students’ learning.

During the computing laboratory sessions, students needed to accomplish the learning tasks. The tasks were designed to promote an exchange of views, sharing of knowledge and resolution of problems that should cultivate a higher level of involvement within a collaborating group. It was intended that communication should play a significant role in a shared activity. The students needed to communicate in order to develop an understanding of the data; to raise questions associated with statistical knowledge; to collaborate on solutions to statistical graphing and computation problems; and to negotiate conflicting views. For instance, a collaborating group would determine how to perform statistical tasks and program Excel. It was anticipated

**Table 1.** Levels of statistical literacy, levels of statistical graphing and the scope of test questions

Questions	Scope	Level	Statistical Literacy	Cook and Weisburg's Model (1999)
1	Data context	One	Comprehension	-
2	Data measurement	One	Comprehension	-
3	Scatterplot construction	Two	Planning and execution	Graph construction
4	Correlation comprehension	Two	Planning and execution	Graph characterization
5	Correlation calculation	Two	Planning and execution	-
6	Correlation test	Two	Planning and execution	-
7	Correlation deduction	Three	Evaluation	Graph inference

that students would gain from collaboration and subsequently develop strategies of problem-solving so as ultimately to have better statistical achievement.

The experimental study reported here was motivated by the need to understand how and to what extent students improve their statistical thinking and graphing in correlation analysis under the influence of social processes within an IT environment. Use of IT enables students to have a more intuitive feel for the concepts being studied; serves to alleviate students' computational burden; and allows them to implement computer logic.

## RESEARCH METHODOLOGY

### Research Participants

Research participants in the experimental study comprised a whole class of 58 full-time students enrolled in Year 2 of the 3-year Higher Diploma in Applied Statistics and Computing (HDASC) course offered by a vocational education institution in Hong Kong. They all had successfully completed secondary education and had passed the Probability and Statistics module in their Year 1 study which covers the topics of basic probability theory and distributions, graphical presentations of data, confidence interval estimation, and statistical hypothesis testing. They were selected because the module was a prerequisite for studying the topic of correlation analysis that was taught in their Year 2 study. HDASC graduates usually join the statistics workforce.

An experimental study consisting of a test was conducted in the computing laboratory because the research participants ought to use computers to access the data secured and managed by a computer server. In addition, the test environment was very similar to the settings of their regular computing laboratory sessions in which they needed to use Excel spreadsheet to accomplish their learning tasks. As a result, this arrangement would reduce the chance of response variability owing to unfamiliar test environment or settings in the present study.

### Experimental Instrument

A test was designed to evaluate key aspects of students' statistical thinking and graphing in an early stage of regression modelling, and used to gather experimental data. In the test, a set of real-life data published by the Census and Statistics Department of the Hong Kong Special Administrative Region was provided, together with a description of the data,  $y$  = electricity consumption (terajoules),  $x_1$  = air temperature ( $^{\circ}$ C),  $x_2$  = relative humidity (%),  $x_3$  = index of industrial production,  $x_4$  = the number of telephone lines (in thousands),  $x_5$  = composite consumer price index, and  $x_6$  = gas consumption (terajoules). The quantity and scope of data were judged to be within the reach of the research participants' ability.

In addition, seven specific questions were designed to evaluate students' responses to each particular task in a preliminary examination of data process. Question 1 was used to evaluate how much students understood the given data regarding the data context that was essential for choosing appropriate data in regression modelling. Question 2 was used to check how well students justified the reasonableness and meaningfulness of data measurements. Both questions are about comprehension acquired in the first level of statistical literacy. Question 3 was to assess students' knowledge of scatterplot construction and proficiency in using Excel graphing tools. Question 4 focused on an appraisal of students' correlation comprehension. Question 5 appraised students' performance of statistical calculations using Excel spreadsheet. Question 6 checked how well students conducted statistical hypothesis testing and reasoned with testing results. These four questions are related to planning and execution in the second level of statistical literacy. Question 7 aimed at assessing students' ability to reason with correlation results and deduce its practical implications. The remaining question is about an evaluation task in the third level of statistical literacy. Alternatively, the first two questions are equivalent to the task of reasoning about data, the fourth and the sixth questions are similar to the task of reasoning about results, and the last question is consistent with the task of reasoning about conclusions according to Bishop and Talbot (2001). The scope of each question and its corresponding level of statistical literacy and statistical graphing capability are presented in **Table 1**.

In Cook and Weisberg's model (1999), graph construction involves converting numerical data into graphical information by using the mechanics of graphing and utilizing all the graphing tools. Graph characterization refers to gaining a qualitative or quantitative summary of the appearance and the shape of a graph, i.e., patterns, trends, centers, clusters, gaps, outliers, spreads, and variations in data and its important features as well as learning visual information about the relative magnitudes of quantities shown on a graph. Moving one step further, graph inference is employed to deduce implicit meanings; to synthesize statistical ideas; to postulate statistical models; to generate statistical hypotheses; to anticipate what further statistical work needs to be carried out; and to devise an action plan for remedial work when necessary.

## Experimental Procedure

Collaborating groups of two students were randomly assigned to one of three experimental conditions: pre-task discussion, post-task discussion or no discussion. The groups who had discussion with their group members either right at the beginning of the test (i.e., pre-task discussion) or in the middle of the test (i.e., post-task discussion) were classified as experimental groups A and B respectively. Individual students who had worked on their own in the entire test period were in a comparison group, i.e., group C. As one student was absent from this test, the total number of students participating in this test became 57. Of these 57 students, 18 (9 pairs) and 16 (8 pairs) were in groups A and B respectively, whereas 23 in group C worked individually. Owing to a random assignment of experimental conditions, equal number of the groups were not obtained.

There were four stages in this experimental study. In Stage I, all students could spend 15 minutes to read a set of data with real-life context from computers either on a group (i.e., students in group A) or an individual basis (i.e., students in groups B and C). But only group A was allowed to have peer interaction within each collaborating group, during which time they could initiate discussions and generate questions associated with measurement, measurement units, content, and context of the data. In Stage II, all students in groups A, B, and C could spend 30 minutes to attempt Questions 1 – 4 individually.

In Stage III, all students could spend 15 minutes to read a set of data with real-life context from computers either on a group (i.e., students in group B) or an individual basis (i.e., students in groups A and C). But group B was allowed to have peer interaction within each collaborating group, and during this time the students could share what, how, and why they had attempted Questions 1 – 4, so as to refine their thoughts; mediate between their conflicting views and promote their individual understanding. In Stage IV, all students in the three groups could spend 30 minutes to attempt Questions 5 – 7 individually.

## Analysis

The students' responses to each test question could be compared among three groups: A (students had pre-task discussion), B (students had post-task discussion), and C (students had no discussion at all) because these groups were independent. To compare the students' responses to statistical thinking and graphing when working on tests "with" and "without" discussion, it was meaningful to combine groups B and C in order to make a genuine comparison of the students' ability to read regression data (Questions 1 and 2); to construct a scatterplot (Question 3); and to read a scatterplot (Question 4) under these two distinctive experimental conditions, i.e., discussion and no-discussion because students in groups B and C did not have any discussion in Stages I and II. Statistical tests were then conducted (see **Tables 4** and **5**) for test responses to Questions 3 and 4 to compare the proportions of complete and correct responses of students in group A versus groups B and C under the normality assumption and in the fulfilment of the condition, i.e., the number of complete and correct response ( $n$ ) is at least 5 in each of the groups. However,  $n$  is less than 5 in either group in **Table 2** and some groups in **Table 3**, so the statistical test could not be carried out for both Questions 1 and 2.

When comparing the response quality of students when working on the test with pre-task discussion and post-task discussion, similar statistical tests were also conducted to compare the proportions of complete and correct responses of students in group A versus group B (see **Tables 7 – 10**).

To summarize the quality of the students' responses to the graphing tasks when working on the test "with" and "without" peer collaboration and "before" and "after" peer collaboration, it is imperative to assess each individual student's work. This was done by performing a qualitative analysis of students' correlation graphing capability by developing a SOLO taxonomy of correlation graphing capability. This approach was chosen because existing assessment frameworks or instruments used in statistics education (e.g., Bude, 2006; Garfield, 2003; Putt et al., 2000; Watson & Callingham, 2003) were not directly applicable to the present study. Bude merely provided a general assessment framework, Garfield evaluated students' statistical reasoning, Putt et al. focused on assessing students' statistical thinking, and Watson and Callingham assessed statistical literacy of primary and secondary school students. This specialized taxonomy was based on the original SOLO taxonomy of Biggs and Collis (1982), and modified in accordance with cognitive model of correlation comprehension (see Li & Goos, 2011). SOLO scores ranging from 1 to 5 were awarded to the quality of their overall responses, i.e., 1 for prestructural; 2 for unistructural; 3 for multistructural; 4 for relational; and 5 for extended abstract.

The prestructural responses are displayed by students who are able to use an appropriate graphing tool but without utilizing graphic features: titles, labels, scales, axis, and symbols. Those students who may use one of the graphic features in their scatterplots have achieved a unistructural achievement. Students whose scatterplots utilize all the graphic features but treat these as isolated entities and/or unrelated to scattering of data, attain a multistructural level of achievement. Integrating the relationship between the measurement, measurement unit, content, and context of data as well as all the graphic features is regarded as relational level of achievement. In attaining the extended abstract level of achievement, students should be able to deduce the qualitative relationship between two variables as unrelated, positively related or negatively related and reveal whether or not such relationship matches or mismatches with the empirical phenomena.

After releasing the normality assumption, the median SOLO scores achieved by students under different experimental conditions were also contrasted using non-parametric statistics, specifically the Fisher exact test because the experimental conditions are independent. First, the test was performed to show whether there is any difference in the response quality of students when working on the test "with" and "without" peer collaboration (see **Table 6**). Second, another test was conducted to reveal whether there is any difference in the response quality of students when working on the test "before" and "after" peer collaboration (see **Table 11**).

Statistical tests so far conducted were under the prescribed experimental conditions and settings as well as assumptions, such as no effect of group composition, no effect of social relationship within groups, no variation of students' statistical competency

**Table 2.** Students' responses to correlation appraisal (Question 1)

Response Categories	Frequency		
	Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=23
Correct answer with justification based on data context	4 (22.2%)	0 (0.0%)	1 (4.3%)
Correct relationship was reported but not providing justification or adequate grounds based on data context	5 (27.8%)	7 (43.8%)	12 (52.1%)
The answer was reliant on graphing/calculation	7 (38.9%)	7 (43.8%)	8 (34.8%)
The answer was unrelated to data context	0 (0.0%)	1 (6.3%)	0 (0.0%)
Unable to assess the possible relationship between two variables	0 (0.0%)	0 (0.0%)	2 (8.7%)
Unattempted	2 (11.1%)	1 (6.3%)	0 (0.0%)

**Note.**

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

and so forth. The conclusive remarks were made on the basis of the p-value, 0.05 (i.e., the probability of making type I error) that is a commonly adopted and widely accepted practice in educational research. In spite of the remarks, the analysis of the data of students' responses shown in the tables could be conducted by other statistical techniques or by applying a p-value other than 0.05 at the reader's discretion. The test results could only indicate whether there is any difference in the response quality of students when working on the test "with" and "without" peer collaboration and "before" and "after" peer collaboration in this experimental study, rather than generalizing a common phenomenon.

**Validity and Reliability of the Analysis**

Validity refers to the credibility and worth of inferences drawn from the data, and reliability to the replicability of the results. Essentially these issues are concerned with the trustworthiness of the findings, i.e., the extent to which the findings are reasonable and justifiable. In some cases, trustworthiness is enhanced if the analysis can be conducted without the researcher knowing the identity of the students or the group to which they have been assigned. Such an approach was not possible here because the first author was also the teacher (and hence assessor) of the students. The trustworthiness of the qualitative analyses of student responses was supported by two additional techniques. First, the second author was independent, and oversaw and validated the first author's adaptation of the SOLO taxonomy and checked its applicability for analyzing student responses to the test questions. Secondly, sufficient examples of responses in each category are provided to allow readers to analyse the data, either by applying the authors' methods or on their own terms.

**RESULTS**

None of the students in groups A, B, and C withdrew from the study and all students completed the tasks on time in each of the four stages, I, II, III, and IV.

**Students' Understanding of Regression Data (Question 1)**

**Table 2** presents the quality of students' responses to hypothesizing about possible correlation with pairs of variables based on the data context. In group A, 22.2% of students provided correct answers together with adequate grounds. This proportion was higher than students of the other two groups, i.e., 0.0% and 4.3% of students in groups B and C, respectively.

In comparing more closely the quality of work between students within the same collaborating group, their work showed consistency in problem approach, explanations, and arguments to a great extent and only a slight discrepancy in wordings or wording sequence was found. It was believed that each of the students was under the influence of discussion with their learning partner.

In addition, 38.9% of students in group A demonstrated a correlation appraisal that was reliant on statistical calculation and/or graphing, compared with 43.8% and 34.8% of students in groups B and C. Five (27.8%), seven (43.8%), and twelve (52.1%) of students in groups A, B, and C group gave correct answers but did not provide justification or adequate grounds based on data context respectively.

**Students' Judgement of data Reasonableness and Meaningfulness (Question 2)**

**Table 3** indicates how well students justified the values of given data covered a reasonable and meaningful range with respect to its context, measurement, and measurement units. Only 11.1%, 0.0% and 21.7% of students in groups A, B, and C respectively could justify the reasonableness and meaningfulness of data measurement with correct and thorough answers, whereas 50.0%, 25.0%, and 43.5% of students in these three respective groups provided correct and valid but incomplete answers. Among these three groups, students of group C exhibited the highest proportion of providing correct and thorough answers, whereas students of group A exhibited the highest proportion of providing correct and valid but incomplete answers.

**Table 3.** Students' judgement of data reasonableness and meaningfulness (Question 2)

Response Categories	Frequency		
	Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=23
Correct and complete answer	2 (11.1%)	0 (0.0%)	5 (21.7%)
Correct answer with partial reasons for meaningful range but nothing for reasonable range	9 (50.0%)	4 (25.0%)	10 (43.5%)
Correct answer but without giving specific/relevant/explicit/valid justification	2 (11.1%)	4 (25.0%)	3 (13.0%)
Correct answer but without any reasons	2 (11.1%)	4 (25.0%)	2 (8.7%)
Unable to assess the possible relationship between two variables	1 (5.6%)	0 (0.0%)	1 (4.3%)
Unattempted	2 (11.1%)	4 (25.0%)	2 (8.7%)

**Note.**

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

**Table 4.** Students' achievement of construction of scatterplot (Question 3)

Response Categories	Frequency		
	Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=22 <sup>b</sup>
Excel graphing tools and syntax were correctly used	13 (72.2%)	6 (37.5%)	11 (50.0%)
Excel graphing tools were correctly used but distracting graph background	1 (5.6%)	0 (0.0%)	0 (0.0%)
Excel graphing tools and syntax were correctly used but the measurement units/axis labels were missing	1 (5.6%)	5 (31.3%)	6 (27.3%)
Excel graphing tools were correctly used but improper graph orientation or graph scales	1 (5.6%)	3 (18.8%)	1 (4.5%)
Excel graphing tools were correctly used but at least two of these mistakes: improper graph orientation, inappropriate graph scales, distracting graph background, and omission of measurement units/axis labels/graph title	2 (11.1%)	2 (12.5%)	4 (18.2%)

**Note.**

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

<sup>b</sup>One student was excluded from this analysis because her computer file was not available.

**Students' Ability to Construct a Scatterplot (Question 3)**

As can be seen in **Table 4**, 72.2% of students in group A demonstrated good knowledge of correlation graphing and proficiency in using Excel graphing tools, compared with 37.5% and 50.0% of students in groups B and C respectively. Among these three groups, students of group A exhibited the highest proportion of correct responses.

There was marginally significant difference in students' ability to construct a scatterplot between "with" and "without" discussion ( $z = 1.9261$ ,  $p = 0.0541$ ). Thus, it is arguable as to whether collaborative discussion might be beneficial to students' later performance on individual tests.

Noticeably, a lower proportion (5.6% versus either 31.3% or 27.3%) of students in group A than students in groups B and C correctly used graphing tools and syntax to construct a scatterplot but without providing measurement units or axis labels. Seemingly, students in groups B and C had worse graphing capability than students in group A in terms of an omission of measurement units as well as axis labels only. This finding cannot be projected onto the other technical mistakes like improper graph orientation, inappropriate graph scales as well as omissions of axis labels, measurement units, and graph title because the proportions of students in any of these three groups who had the other technical mistakes were more or less the same.

**Students' Responses to Reading a Scatterplot (Question 4)**

The quality of students' responses to reading the scatterplot is presented in **Table 5** which reveals 33.3%, 18.8% and 17.4% of students in groups A, B, and C respectively could comprehend correlation patterns in scatterplots with valid reasons. Students of group A exhibited the highest proportion of reading correlation patterns but did not show better responses to reading scatterplots than groups B and C ( $z = 1.2868$ ,  $p = 0.1982$ ).

Among those students who could comprehend correlation patterns in scatterplots but did not provide any reasons, group A had the highest proportion (33.3% versus 18.8% and 4.3% of students in groups B and C). In addition, 11.1%, 6.3% and 34.8% of students in groups A, B, and C respectively did not attempt this question. Among them, group C had the highest non-response rate. Likewise, 5.6%, 25.0% and 17.4% of students in the three groups gave incorrect or imprecise answers to this question respectively. Their incorrect answers were due to inappropriate graph scales or wrong reasons. They had given imprecise answers as they provided inexplicit explanations or reasons irrelevant to data scattering. Among them, group A had the lowest proportion.

To sum up, students who had discussion demonstrated higher ability in reading scatterplots. Such reading required making connections between correlation patterns in the scatterplots and the data context (see Li & Goos, 2011). The connection tasks would be facilitated by discussion in which students checked the spatial association between pairs of data (i.e.,  $x_i, y_i$ ) physically located on a scatterplot as well as the global trend of spatial representation of pairs of data. This piece of information was displayed on the computer monitor as visible parts of problem-solving products which could be used to substantiate each individual student's claim or assisted in development of shared understanding between students too.

**Table 5.** Students' responses to reading scatterplot (Question 4)

Response Categories	Frequency		
	Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=23
Correct answers with valid reasons	6 (33.3%)	3 (18.8%)	4 (17.4%)
Correct answers but without any reasons	6 (33.3%)	3 (18.8%)	1 (4.3%)
Correct answers but incomplete or conflicting reasons	2 (11.1%)	5 (31.3%)	5 (21.7%)
Correct answers reliant on calculations	1 (5.6%)	0 (0.0%)	0 (0.0%)
Incorrect/imprecise answers	1 (5.6%)	4 (25.0%)	4 (17.4%)
Unable to estimate the correlation coefficient	0 (0.0%)	0 (0.0%)	1 (4.3%)
Unattempted	2 (11.1%)	1 (6.3%)	8 (34.8%)

**Note.**

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

**Table 6.** Frequency distribution of students' SOLO scores of graph construction and graph characterization

SOLO score	SOLO descriptions	Frequency		
		Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=22 <sup>b</sup>
1	Prestructural	0 (0.0%)	0 (0.0%)	0 (0.0%)
2	Unistructural	5 (27.8%)	10 (62.5%)	11 (50.0%)
3	Multistructural	6 (33.3%)	5 (31.3%)	10 (45.5%)
4	Relational	7 (38.9%)	1 (6.3%)	1 (4.5%)
Median <sup>c</sup>		3	2	2.5

**Notes.**

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

<sup>b</sup>One student was excluded from this analysis because her computer file was not available.

<sup>c</sup>A 4-point scale based on student's responses to correlation graphing.

Responses to Questions 3 and 4 were subjected to further analysis to evaluate students' graph construction and graph characterization using the SOLO taxonomy. The construction and characterization are the first two phases of creating and using statistical graphs outlined by Cook and Weisberg (1999). **Table 6** presents a frequency distribution of students' SOLO scores and the medians for SOLO scores in the three different groups. A single SOLO score, 1 (prestructural); 2 (unistructural); 3 (multistructural); or 4 (relational) was awarded to the quality of responses each student gave to Questions 3 and 4. It should be noted that 5 (extended abstract) would not be awarded as these questions did not ask students to deduce the practical implications of data relationship.

**Table 6** shows that 0.0% (none) of students in these three groups gave prestructural responses because they could construct scatterplots together with at least one of these graphic features: titles, axis-labels, and axis-scales. In group A, 38.9% of students gave relational responses, indicating that they integrated the relationship between measurement, measurement units, content and context of data, whereas only 6.3% and 4.5% of students in both groups B and C recorded this type of response. Among all these proportions, group A had higher capability of graph construction and graph characterization. It seemed that students who had pre-task discussion demonstrated higher capability of graph construction and graph characterization than those who had post-task or no discussion.

It should be noted that students in group B had no discussion prior to attempting these questions so that groups B and C should therefore be combined in order to make a genuine comparison of the quality of students' correlation graphing under these two distinctive experimental conditions, i.e., discussion and no-discussion. The Fisher exact test ( $p = 0.003$ ) showed that there was a significant difference in students' median SOLO scores between "with" and "without" discussion. Discussion could therefore account for group A achieving a higher median SOLO score, indicating that in general students' responses displayed multistructural features in terms of graph construction and graph characterization. With discussion, then, it appears that students could identify and utilize all the graphic features to construct scatterplots (i.e., multistructural) but might not fully integrate the relationship between the measurement, measurement units, content and context of data, as well as all the graphic features (i.e., relational). Summing up, students who had discussion generally showed better quality of graph construction and graph characterization.

**Students' Responses to Correlation Calculation (Question 5)**

As can be seen in **Table 7**, 88.9%, 62.5% and 60.9% of students in groups A, B, and C respectively used Excel tools to accomplish correlation calculation tasks including proper selection and use of correlation function or correlation analysis tool and correct input of data and output of correlation results. Students of group A exhibited the highest proportion of accomplishing correlation calculation tasks.



**Table 7.** Students' responses to correlation calculation (Question 5)

Response Categories	Frequency		
	Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=23
Correct answer, Excel tool, and syntax	16 (88.9%)	10 (62.5%)	14 (60.9%)
Correct Excel tool and syntax were used but no implications	1 (5.6%)	1 (6.3%)	2 (8.7%)
Correct answers but unable to assess student's Excel proficiency because computer file was unavailable	1 (5.6%)	3 (18.8%)	6 (26.1%)
Unable to assess student's Excel proficiency because computer file was unavailable	0 (0.0%)	1 (6.3%)	1 (4.3%)
Incorrect Excel tool and syntax were used	0 (0.0%)	1 (6.3%)	0 (0.0%)

Note.

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

**Table 8.** Students' knowledge of Excel programming and syntax (Question 6)

Response Categories	Frequency		
	Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=23
Correct Excel programming	15 (83.3%)	9 (56.3%)	12 (52.2%)
Incorrect Excel syntax/programming	1 (5.6%)	1 (6.3%)	1 (4.3%)
Unable to assess Excel programming because computer file was unavailable	2 (11.1%)	5 (31.3%)	10 (43.5%)
Unattempted	0 (0.0%)	1 (6.3%)	0 (0.0%)

Note.

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

A statistical test was performed to study whether students who engaged in post-task discussion showed better ability to use Excel calculation than those who had pre-task discussion. The difference in proportions between them (i.e., group A versus group B) was not statistically significant ( $z = 1.8106$ ,  $p = 0.0702$ ) but arguable. Students who engaged in post-task discussion possibly showed worse ability to use Excel calculation tools than the students who engaged in pre-task discussion.

#### Students' Knowledge of Excel Programming and Syntax (Question 6)

Students' responses to Question 6 were evaluated based on two criteria. The first criterion dealt with students' knowledge of Excel syntax and programming skills as well as the second with their performance of statistical hypothesis testing. Programming Excel is about to build an executable program to perform a statistical hypothesis test. **Table 8** shows that 83.3%, 56.3% and 52.2% of students in groups A, B, and C respectively programmed Excel properly for statistical hypothesis testing. These proportions of sound knowledge of Excel syntax and good Excel programming skills possessed by the students were substantially high. However, it was not possible to assess Excel programming for 11.1%, 31.3% and 43.5% of students in groups A, B, and C respectively because computer files were unavailable.

To investigate whether students of group A outperformed students of group B, it was found that students who engaged in pre-task discussion showed no better performance of correlation calculation using Excel than those who had post-task discussion ( $z = 1.7299$ ,  $p = 0.0836$ ). Nevertheless, it should be noted that a substantial proportion of these students' Excel programming proficiency could not be assessed because their computer files were unavailable.

In addition, substantially low proportions (i.e., only one in each of the three groups) of students had used incorrect Excel syntax or programmed Excel incorrectly. For example, a parenthesis was misplaced in an Excel function or the number of paired data ( $n$ ) was mis-counted and varying data count was encountered. The overall findings did not support the argument that students who had discussion, regardless of the time of discussion being held, showed better proficiency in Excel programming.

#### Students' Performance of Statistical Hypothesis Testing (Question 6)

Students' responses to Question 6 were then evaluated to compare how well they performed statistical hypothesis testing. It was found that 16.7%, 25.0% and 39.1% of students in groups A, B, and C respectively accomplished statistical hypothesis testing tasks in which they provided proper formulation of null and alternative hypotheses; correct statistical evidence and decision; sound reasoning with statistical evidence from Excel output as well as statistical implications (**Table 9**).

**Table 9.** Students' performance of statistical hypothesis testing (Question 6)

Response Categories	Frequency		
	Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=23
Correct and complete answers	3 (16.7%)	4 (25.0%)	9 (39.1%)
Correct answers but giving no/inadequate statistical evidence	0 (0.0%)	0 (0.0%)	2 (8.7%)
Correct answers but giving no/incorrect implications	9 (50.0%)	6 (37.5%)	0 (0.0%)
No/incorrect rejection region was given	4 (22.2%)	2 (12.5%)	7 (30.4%)
No statistical decision was made	1 (5.6%)	1 (6.3%)	1 (4.3%)
Wrong statistical tools/tests	0 (0.0%)	1 (6.3%)	4 (17.4%)
Unattempted	1 (5.6%)	2 (12.5%)	0 (0.0%)

**Note.**

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

**Table 10.** Students' correlation deduction and synthesis (Question 7)

Response Categories	Frequency		
	Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=23
Correct data relationship	3 (16.7%)	0 (0.0%)	0 (0.0%)
Correct strength and/or direction of data relationship but without deducing the relationship in practical context	8 (44.4%)	13 (81.3%)	16 (69.6%)
Unable to synthesize data relationship from Q6 but matched with results from Q1	2 (11.1%)	2 (12.5%)	3 (13.0%)
Incorrect strength and/or direction of data relationship and no deduction	2 (11.1%)	0 (0.0%)	0 (0.0%)
Unrelated matters were highlighted	0 (0.0%)	0 (0.0%)	1 (4.3%)
Unattempted	3 (16.7%)	1 (6.3%)	3 (13.0%)

**Note.**

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

As the proportions of students in any of these three groups who successfully accomplished statistical hypothesis testing tasks were substantially low, it would be meaningful to compare how poorly they did in the test held in Stage IV: fourteen (77.8%), ten (62.5%), and fourteen (60.9%) of students in groups A, B, and C respectively failed to complete statistical hypothesis testing tasks. Each of these percentages is the sum of the entries associated with their failures in each respective group that were due to no/inadequate statistical evidence; no/inadequate implications for correlation test results; no/incorrect rejection region; no statistical decisions made; or wrong statistical tools/tests used. When comparing two experimental conditions, pre-task discussion (group A) versus post-task discussion (group B), the difference between these proportions of incorrect responses (i.e., 77.8% vs 62.5%) was not statistically significant ( $z = 0.9759$ ,  $p = 0.3291$ ), implying that students in any of these two groups did poorly in the test held in Stage IV irrespective of types of students' mistakes; incompetency; or experimental conditions.

A relatively higher proportion (17.4% versus either 0.0% or 6.3%) of students in group C than students in both groups A and B adopted wrong statistical tools or tests. Specifically, students made a direct comparison of a correlation coefficient with a z-value (standard normal deviate) for performing statistical hypothesis testing. Obviously, students did not give correct rejection region owing to using incorrect probability distribution; misreading z-value from Excel statistical function; mixing up the rationales of one-sided and two-sided tests, particularly without stating null and alternative hypotheses; or wrong Excel programming. Inappropriate statistical tests or wrong statistical decisions resulted from these technical mistakes and eventually led to drawing an inconsistent conclusion or a wrong implication.

**Students' Correlation Reasoning (Question 7)**

The quality of students' responses to interpreting correlation results and deducing its practical implications are summarized in **Table 10**. A higher proportion (81.3% versus 44.4% or 69.6%) of students in group B than groups A and C responded to correlation deduction and synthesis vaguely and their arguments were not linked to the data context, so it is worth noting that 16.7% of students in group A could deduce a data relationship in a practical context but no student (0.0%) in the other two groups was successful.

To summarize, Questions 3, 4, 6, and 7 formed the basis of evaluating students' overall responses in preliminary examination of data using the SOLO taxonomy, focusing on graph construction, graph characterization, and graph inference. In **Table 11**, 0.0% (none) of students in these three groups gave pre-structural responses. Only three students in group A (16.7%) scored 5, indicating that they could deduce correlation between two variables, whereas 0.0% of students in both groups B and C were able to do this, as the proportions of students in any of these three groups who made at least one of these technical mistakes, such as improper graph orientation, inappropriate graph scales, omissions of axis labels and measurement units was high (refer to **Table 4**). Any of these technical mistakes was serious and could prevent students from deducing the qualitative relationship between two variables. Obviously, improper graph orientation exchanged an independent variable ( $x$ ) and a dependent ( $y$ ) variable so that students got confused and subsequently misconceived of the data relationship, i.e.,  $x$  became a function of  $y$ . Inappropriate graph scales distorted the pattern on a scatterplot and consequently led them to mis-appraise correlation from a scatterplot (e.g., Sun et al., 2016). An omission of axis labels misled students to treat graphic features as isolated entities and/or unrelated to a

**Table 11.** Frequency distribution of students' SOLO scores of preliminary examination of data

SOLO score	SOLO descriptions	Frequency		
		Group A <sup>a</sup> N=18	Group B <sup>a</sup> N=16	Group C <sup>a</sup> N=22 <sup>b</sup>
1	Prestructural	0 (0.0%)	0 (0.0%)	0 (0.0%)
2	Unistructural	5 (27.8%)	10 (62.5%)	11 (50.0%)
3	Multistructural	6 (33.3%)	5 (31.3%)	10 (45.5%)
4	Relational	4 (22.2%)	1 (6.3%)	1 (4.5%)
5	Extended abstract	3 (16.7%)	0 (0.0%)	0 (0.0%)
Median <sup>a</sup>		3	2	2.5

**Notes.**

<sup>a</sup>Collaborating groups of students who had discussion with their group members either right at the beginning of the test or in the middle of test were classified as groups A and B respectively. Individual students who had worked on their own in the entire test period were in group C.

<sup>b</sup>One student was excluded from this analysis because her computer file was not available.

<sup>c</sup>A 5-point scale based on student's responses to correlation graphing.

correlation pattern. An omission of measurement units concealed the physical meanings and magnitude of data. Nevertheless, 22.2%, 6.3% and 4.5% of students in these respective groups scored 4, illustrating that they could integrate the relationship between the measurement, measurement unit, content, and context of data. Among all these proportions, group A demonstrated the highest quality in preliminary examination of data, thus supporting the benefits of pre-task discussion.

Among the three different groups, students of group A achieved the highest median SOLO score (3 versus 2 and 2.5). In other words, students who had engaged in pre-task discussion, achieved better quality in correlation graphing tasks. Students who had post-task discussion, gave responses of somewhat similar quality to those who had no discussion in the entire test period.

To compare changes in the quality of student responses "before" and "after" peer collaboration by using the Fisher exact test, the difference in median SOLO scores between these two groups was statistically significant ( $p = 0.0280$ ), implying that students who had pre-task discussion had higher SOLO achievement than those who had post-task discussion.

Apparently, students who experienced pre-task discussion displayed a higher level of correlation graphing that suggests they would subsequently progress in regression modelling if they could give explanations to support their correlation appraisal. Giving explanations may have prompted students to articulate and reflect on their thinking as well as regulate their insights. In particular, students may have shared statistical results to help each other in order to internalize and regulate strategies for successful regression modelling when working in collaborating groups.

## LIMITATIONS

Some caution is needed in interpreting results of data analysis for three reasons. First, the experimental study reported the achievement outcomes between three experimental conditions, pre-task discussion, post-task discussion, and no discussion, but the degree of involvement in collaborating groups for these three conditions could not be contrasted. The conclusions were drawn on the basis of analyses of research participants' written answers and graphing tasks in a given test. But, how actual social processes within collaborating groups took place in these experimental conditions was not clearly known because neither audiotaping nor videotaping could be carried out to take detailed account of social processes.

Second, the findings of the present study, which is confined to the topic of correlation analysis, cannot be generalized to peer collaboration taking place in any kind of statistics learning because the approach to statistical thinking and graphing can vary for different statistical topics. Third, the results were obtained under the prescribed experimental conditions and settings as well as assumptions. Statistical analysis of data was performed at the 5% level of significance to derive summative results which could only indicate the likelihood of statistical significance, rather than generalizing definite answers or a common phenomenon.

## CONCLUSION

With pre-task discussion, students (group A) achieved a higher level of understanding of regression data than those who did not have discussion at this time (both groups B and C). The students could hypothesize about possible correlation with pairs of variables based on the data context (Question 1). For example, the students should be all be aware of one common phenomenon in Hong Kong. That is, most households had air conditioners but not heaters and they generally turned on air conditioners in hot weather. For those households that did have heaters, they might not turn on their heaters in winter because they found the winter in Hong Kong was not cold enough. Of course, students might say there was no relationship between the electricity consumption and air temperature if they could substantiate their answer on the assumption that many households might turn on their heaters in winter. The quality of students' responses was evaluated according to how well they made connection among facts or evidence and deduced the relationship between them if any.

The students in group A showed better quality in graph construction as well as graph characterization. First, they constructed better scatterplots with fewer graphing mistakes, such as graphing tools and syntax, graph title, graph orientation, graph scales, measurement units, and axis labels (Question 3). Second, they demonstrated higher ability in reading scatterplots. When reading scatterplots (Question 4), the students adopted Li and Goos's (2011) model to check patterns of data on a scatterplot; then to

retrieve or construct qualitative and quantitative meanings from the data pattern; and eventually to relate the meanings to the titles, labels, and scales in the scatterplot. Unfortunately, the influence of the pre-task discussion on a preliminary examination of regression data was not explicit in accomplishing tasks of judgement of data reasonableness and meaningfulness (Question 2). Correct and complete responses to these four questions required broader knowledge and deeper thinking for the tasks of reasoning about data and reasoning about results. The reasoning tasks demanded for making sense of the data connecting with their individual contexts, thus justifying and blending multiple perspectives or divergent views. The emergence of such perspectives or views seems less likely when students are working alone than in collaboration with peers. Communication is evident in the collaboration such that they make their ideas available for comment, suggestion, and argument, thus developing reasoning within the social process of learning.

Theoretically speaking, students in group B should perform the tasks of correlation calculation, Excel programming, statistical hypothesis testing, and reasoning as well as students in group A because both groups were allowed to have peer collaboration within each collaborating group, although discussions were held in different stages of the test. Group A had discussion right at the beginning of the test (i.e., Stage I), whereas group B had discussion the middle of the test (i.e., Stage III) respectively. Nevertheless, it was found that a substantial proportion of the student responses to Questions 5 - 7 presented evidence of improvement after pre-task discussion which might form the basis for accomplishing tasks from abstract to concrete.

After participating in pre-task discussion, students also had meaningful gains in computing ability, Excel programming proficiency, and statistical hypothesis testing performance when attempting Questions 5 and 6. The gains may be more or less related to articulation, self-evaluation, and re-organization of thoughts. The articulation of thoughts is about using Excel tools and syntax. Self-evaluation concerns programming logic, the setup of the null versus alternative statistical hypotheses, the selection of one-sided versus two-sided test, and the appropriateness of the probability distribution (normal distribution versus t-distribution). Re-organization of thoughts would be necessary if students found the logic inconsistent arising from discrepancies among the hypotheses setup, the test selection, and the probability distribution because these are interrelated. Furthermore, the students also did better in correlation reasoning (Question 7), which necessitates verbal thoughts for deducing the practical implications of correlation results by consolidating what they had learnt from the data, the scatterplots they constructed, and also the statistical hypothesis testing they carried out, probably under the influence of pre-task discussion.

More interestingly, students who had pre-task discussion demonstrated sound understanding of data and gave more complex and connected responses (SOLO analysis), whereas students who had post-task or no discussion gave responses with lower SOLO scores. Three possible factors may have contributed to students' SOLO scores and the improved quality of students' responses to Questions 5 - 7: their own understanding of the data prior to discussion, the quality of group discussion and the nature of students' participation and involvement in their group discussion.

The linkage between social processes and correlation analysis tasks in an IT environment could be understood from a socio-cultural perspective by focusing on the issue of whether or not students having discussion would tend to have better correlation graphing. The overall findings of how students benefited from group discussion were consistent with socio-cultural theory of Vygotsky (1978) in which knowledge construction necessarily promotes different ways of thinking and incorporates different perspectives of their learning partners.

## FUTURE RESEARCH DIRECTIONS

The present study leaves two broad questions. Although this experimental study could merely contrast the quality of student responses in a test "with" and "without" peer collaboration; as well as "before" and "after" peer collaboration, it did not show how peer collaboration took place. For this reason, it is recommended to conduct another experimental study during which students' verbal speech and their social interaction when doing correlation analysis tasks at computers, such as keying in data; programming Excel; reading the screen displays of computer output; and so forth are videotaped at the same time. Analyzing the videotaped data should provide an account of how students' development of statistical thinking and proficiency in statistical graphing benefit from their verbal exchanges and social interaction in an IT environment.

Second, the task of reasoning about data was not well exhibited by the research participants (HDASC students). This task also demands statistical communication in written form associated with the use of the English language, which is a vital part of the HDASC course aiming at equipping students with knowledge and skills for their prospective careers in statistics. However, students were unable to accomplish the task, probably owing to the language of statistics being a hurdle most students cannot get over (see Dunn et al., 2016), or poor English proficiency. Although Dunn et al. proposed various ways of enhancing students' statistical communication skills, the effectiveness of these approaches need to be evaluated under real classroom conditions.

**Author contributions:** All authors have sufficiently contributed to the study, and agreed with the results and conclusions.

**Funding:** No funding source is reported for this study.

**Declaration of interest:** No conflict of interest is declared by authors.

## REFERENCES

- Biehler, R. (2016). Professional development for teaching probability and inferential statistics with digital tools at upper secondary level. *Proceedings of the 13<sup>th</sup> International Congress on Mathematical Education*, Hamburg, Germany: International Commission on Mathematical Education. [http://iase-web.org/documents/papers/icme13/ICME13\\_11\\_Biehler.pdf](http://iase-web.org/documents/papers/icme13/ICME13_11_Biehler.pdf)

- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press.
- Bishop, G., & Talbot, M. (2001). Statistical thinking for novice researchers in biological science. In C. Batanero (Ed.), *Proceedings of the IASE Round Table Conference on Training Researchers in the Use of Statistics* (pp. 215-226).
- Bude, L. (2006). Assessing students' understanding of statistics. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7<sup>th</sup> International Conference on Teaching Statistics*. Bahia, Brazil: International Association for Statistical Education. [http://iase-web.org/documents/papers/icots7/6G3\\_BUDE.pdf](http://iase-web.org/documents/papers/icots7/6G3_BUDE.pdf)
- Carter, H. & Milne, C. (2000). Moving online: Developing a compulsory university-wide statistical literacy program. In M. Wallace, A. Ellis, & D. Newton (Eds.), *Proceedings of Moving Online: A Conference to Explore the Challenges for Workplaces, Colleges and Universities* (pp. 155-162). Gold Coast, Australia.
- Casey, S.A. & Wasserman, N.H. (2015). Teachers' knowledge about informal line of best fit. *Statistics Education Research Journal* 14(1), 8-35.
- Cook, R. D., & Weisberg, S. (1999). Graphs in statistical analysis: Is the medium the message? *The American Statistician*, 53(1), 29-37. <https://doi.org/10.1080/00031305.1999.10474426>
- Curran, E., Carlson, K., & Celotta, D.T. (2013). Changing attitudes and facilitating understanding in the undergraduate statistics classroom: A collaborative learning approach. *Journal of the Scholarship of Teaching and Learning*, 13(2), 49-71.
- da Silva, C. B., Kataoka, V. Y., & Cazorla, I. (2014). Analysis of teachers' understanding of variation in the dot-boxplot context. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the 9<sup>th</sup> International Conference on Teaching Statistics*, Arizona, USA: International Association for Statistical Education. [http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9\\_3A2\\_SILVA.pdf?1405041597](http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_3A2_SILVA.pdf?1405041597)
- Dunn, P. K., Carey, M. D., Richardson A. M., & McDonald, C. (2016). Learning the language of statistics: Challenges and teaching and approaches. *Statistics Education Research Journal*, 15(1), 8-27.
- Fernandez, M. S., Pomilio, C., Cueto, G., Filloy, J., Gonzalez-Arzac, A., Lois-Milevicich, J., & Perez, A. (2020). Improving skills to teach statistics in secondary school through activity-based workshops. *Statistics Education Research Journal*, 19(1), 106-119.
- Gal, I. (2003). Expanding conceptions of statistical literacy: An analysis of products from statistics agencies. *Statistics Education Research Journal*, 2(1), 3-21. <https://doi.org/10.1198/0003130031469>
- Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38.
- Garfield, J. (2011). Statistical literacy, reasoning, and thinking. In M. Lovric (Ed.), *International Encyclopedia of Statistical Science* (pp. 1439-1442). Springer. [https://doi.org/10.1007/978-3-642-04898-2\\_81](https://doi.org/10.1007/978-3-642-04898-2_81)
- Goos, M. (2004). Learning Mathematics in a classroom community of inquiry. *Journal for Research in Mathematics Education*, 35(4), 258-291. <https://doi.org/10.2307/30034810>
- Goos, M. (2009). A sociocultural framework for understanding technology integration in secondary school mathematics. In M. Tzekaki, M. Kaldrimidou, & H. Sakonidis (Eds.), *Proceedings of the 33<sup>rd</sup> Conference of the International Group for the Psychology of Mathematics Education* (vol. 1, pp. 113-120), Psychology of Mathematics Education.
- Goos, M. (2014). Technology integration in secondary school mathematics: The development of teachers' professional identities. In A. Clark-Wilson, O. Robutti, & N. Sinclair (Eds.), *The Mathematics Teacher in the Digital Era: An International Perspective on Technology Focused Professional Development* (pp. 139-161), Springer. [https://doi.org/10.1007/978-94-007-4638-1\\_7](https://doi.org/10.1007/978-94-007-4638-1_7)
- Gordon, I., & Finch, S. (2010). How we can all learn to think critically about data. In C. Reading (Ed.), *Proceedings of the 8<sup>th</sup> International Conference on Teaching Statistics*, Ljubljana, Slovenia: International Association for Statistical Education. [http://iase-web.org/documents/papers/icots8/ICOTS8\\_7G4\\_GORDON.pdf](http://iase-web.org/documents/papers/icots8/ICOTS8_7G4_GORDON.pdf)
- Grant, R. (2017). Statistical literacy in the data science workplace. *Statistics Education Research Journal*, 16(1), 17-21.
- Haines, B. (2015). Conceptualizing a framework for advanced placement statistics teaching knowledge. *Journal of Statistics Education*, 23(3). <https://doi.org/10.1080/10691898.2015.11889747>
- Horton, N. J., & Hardin, J. S. (2015). Teaching the next generation of statistics students to think with "data": Special issue on statistics and the undergraduate curriculum. *The American Statistician*, 69(4), 259-265. <https://doi.org/10.1080/00031305.2015.1094283>
- Li, K. W. K. (2006). *Information technology in the teaching and learning of statistics: Educational issues and practices* (Unpublished doctoral thesis). The University of Queensland, Brisbane.
- Li, K. W., & Goos, M. (2011). Reinforcing students' correlation comprehension. *The International Journal of Learning*, 17(11), 261-274.
- Malaspina, M., & Malaspina, U. (2020). Game invention as means to stimulate probabilistic thinking. *Statistics Education Research Journal*, 19(1), 57-72.
- Mogro-Wilson, C., Reeves, M. G., & Charter, M. L. (2015). Applying sociocultural theory to teaching statistics for doctoral social work students. *Journal of Teaching in Social Work*, 35(1-2), 151-165. <https://doi.org/10.1080/08841233.2014.976354>
- Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning: An International Journal*, 13(1-2), 27-46. <https://doi.org/10.1080/10986065.2011.538302>
- Pfannkuch, M., & Budgett, S. (2016). Markov processes: Exploring the use of dynamic visualizations to enhance student understanding. *Journal of Statistics Education*, 24(2), 63-73. <https://doi.org/10.1080/10691898.2016.1207404>

- Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17-46). Kluwer Academic Publishers. [https://doi.org/10.1007/1-4020-2278-6\\_2](https://doi.org/10.1007/1-4020-2278-6_2)
- Pierce, R., Chick, H., Watson, J., Les, M., & Dalton, M. (2014). A statistical literacy hierarchy for interpreting educational system data. *Australian Journal of Education*, 58(2), 195-217. <https://doi.org/10.1177/0004944114530067>
- Putt, I., Perry, B., Jones, G., Thornton, C., Langrall, C., & Mooney, E. (2000). Primary school students' statistical thinking: A comparison of two Australian states. *Proceedings of the 23<sup>th</sup> Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 519-526). Mathematics Education Research Group of Australasia.
- Reaburn, R. (2017). Statistics instructors' belief and misconceptions about p-values. In A. Downton, S. Livy, & J. Hall (Eds.), *Proceedings of the 40<sup>th</sup> Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 428-433). Mathematics Education Research Group of Australasia.
- Schau, C., Millar, M., & Petocz, P. (2012). Research on attitudes towards statistics. *Statistics Education Research Journal*, 11(2), 2-5.
- Schiold M. (1998). Statistical literacy and evidential statistics. *Proceedings of the Section on Statistical Education of the Annual Meeting of American Statistical Association*, California, USA: American Statistical Association. <http://www.statlit.org/pdf/1998SchioldASA.pdf>
- Schiold M. (1999). Statistical literacy: Thinking critically about statistics. *Of Significance Journal*, 1(1). <http://web.augsburg.edu/~schiold/MiloPapers/984StatisticalLiteracy6.pdf>
- Schiold M. (2000). Statistical literacy and mathematical thinking. *Paper presented at the 9<sup>th</sup> International Congress on Mathematics Education*, Tokyo, Japan. <http://www.statlit.org/PDF/2000SchioldICME.pdf>
- Schiold M. (2001). Statistical literacy: Reading tables of rates and percentages. *Paper presented at the Section on Statistical Education of the Annual Meeting of American Statistical Association*, Atlanta, USA. <http://www.statlit.org/pdf/2001SchioldASA2.pdf>
- Schiold M. (2002). Three kinds of statistical literacy: What should we teach? *Proceedings of the 6<sup>th</sup> International Conference on Teaching Statistics*, Cape Town, South Africa: International Association for Statistical Education. [http://www.iase-web.org/documents/papers/jicots6/1a2\\_schi.pdf](http://www.iase-web.org/documents/papers/jicots6/1a2_schi.pdf)
- Schiold, M. (2010). Assessing statistical literacy: Take CARE. In P. Bidgood, N. Hunt, & F. Jolliffe (Eds.) *Assessment methods in statistical education: An international perspective* (pp. 133-152). Wiley. <https://doi.org/10.1002/9780470710470.ch11>
- Schiold, M. (2014). Odyssey: A journey to lifelong statistical literacy, In K. Makar, B. de Sousa, & R. Gould (Eds.), *Proceedings of the 9<sup>th</sup> International Conference on Teaching Statistics*, Arizona, USA: International Association for Statistical Education. [http://iase-web.org/jicots/9/proceedings/pdfs/ICOTS9\\_7A1\\_SCHIELD.pdf?1405041685](http://iase-web.org/jicots/9/proceedings/pdfs/ICOTS9_7A1_SCHIELD.pdf?1405041685)
- Sharma, S. (2014). Influence of culture on secondary school students' understanding of statistics: A Fijian perspective. *Statistics Education Research Journal*, 13(2), 104-117.
- Sotos, A. E. C., Vanhoof, S., Noortgate, W. V. D., & Onghena, P. (2009). The transitivity misconception of Pearson's correlation coefficient. *Statistics Education Research Journal*, 8(2), 33-55.
- Sun, Y., Li, S., Bonini, N., & Liu, Y. (2016). Effect of graph scale on risky choice: Evidence from preference and process in decision-making, *PLoS One*, 11(1), 1-12. <https://doi.org/10.1371/journal.pone.0146914>
- Valsiner, J. (1997). *Culture and the development of children's action: A theory of human development* (2<sup>nd</sup> edition). Wiley.
- Vygotsky, L. S. (1978). *Mind in society*. Harvard University Press.
- Wang, L., Bruce, C. S., & Hughes, H. (2011). Sociocultural theories and their application in information literacy research and education. *Australian Academic and Research Libraries*, 42(4), 296-308. <https://doi.org/10.1080/00048623.2011.10722242>
- Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Lawrence Erlbaum Associates, Inc., Publishers.
- Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3-46.
- Watson, J., & Neal, D. (2012). Preparing students for decision-making in the 21<sup>st</sup> century – Statistics and probability in the Australian curriculum: mathematics. In B. Atweh, M. Goos, R. Jorgensen, & D. Siemon (Eds.), *Engaging the Australian curriculum mathematics – Perspectives from the field* (pp. 89-113). Mathematics Education Research Group of Australasia.
- Wild, C., Pfannkuch, M., Regan, M., & Parsonage, R. (2015). Accessible conceptions of statistical inference: Pulling ourselves up by the Bootstraps. *International Statistical Review*, 85(1), 84-107. <https://doi.org/10.1111/insr.12117>