J. Korean Soc. Math. Ed. Ser. D. (2023) 26(3), 127–146
https://doi.org/10.7468/jksmed.2023.26.3.127

ISSN 1226-6191
Online ISSN 2287-9943

RESEARCH ARTICLE

Fostering Students' Statistical Thinking through Data Modelling

Ken W. Li[1]

[1]Retired Scholar

Received: February 10, 2023 / Accepted: April 21, 2023 / Published online: September 28, 2023

Abstract

Statistical thinking has a broad definition but focuses on the context of regression modelling in the present study. To foster students' statistical thinking within the context, teaching should no longer be seen as transfer of knowledge from teacher to students but as a process of engaging with learning activities in which they develop ownership of knowledge. This study aims at collaborative learning contexts; students were divided into small groups in order to increase opportunities for peer collaboration. Each group of students was asked to do a regression project after class. Through doing the project, they learnt to organize and connect previously accrued piecemeal statistical knowledge in an integrated manner. They could also clarify misunderstandings and solve problems through verbal exchanges among themselves. They gave a clear and lucid account of the model they had built and showed collaborative interactions when presenting their projects in front of class. A survey was conducted to solicit their feedback on how peer collaboration would facilitate learning of statistics. Almost all students found their interaction with their peers productive; they focused on the development of statistical thinking with concerted effort.

Keywords: collaborative learning, project work, regression model, vocational education

• email: dr_ken_li@yahoo.com.hk

I. INTRODUCTION

There is a need for developing students' statistical thinking to prepare them for decision-making in the 21st Century so as to make statistical sense of our world (Bargagliotti et al., 2020; Frischemeier et al., 2022; Watson & Neal, 2012). The type of thinking is so acquired for evaluating information critically, and appreciating statistical evidence. Unfortunately, statistical thinking is not well grasped by many students, for instance, a statistical term, "correlation" was commonly misconceived or misinterpreted as being causality and having transitivity (Sotos et al., 2009). On other hand, statistical thinking does not have a general definition. Statistical thinking can also be defined as narrowly as focusing on the study of variation, whereas the definition of statistical thinking can be as broad as the thought processes involved in statistical activities or works.

Bishop and Talbot (2001) defined statistical thinking as the thought processes involved in reasoning about data; reasoning about results; and reasoning about conclusions. Reasoning about data involves studying the background, content, and context of data and justifying reasonableness of data measurement, appropriateness of measurement units of data, and meaningfulness of the data range. Reasoning about results evaluates whether outputs resulting from statistical processes are consistent with statistical logic and any have interconnection among or discrepancy between them. When reasoning about conclusions, one draws valid conclusions based on statistical evidence.

Hoerl and Snee (2010) pointed out that understanding process variation is at the heart of statistical thinking, and there were two kinds of variation. First, the variation produced by the processes that create products and services. Second, the variation produced by the measurement processes that generate the data we use to characterize the performance of our processes. Mallows (1998) and Moore (1997) supplemented Hoerl and Snee's definition by adding the component, "reasoning about data".

Similarly, Dransfield et al. (1999) argued that process and system thinking were elements of statistical thinking and outlined a hierarchy of statistical thinking as: strategic, tactical, and operational levels. The strategic level of statistical thinking involves devising a scheme or an overall plan for achieving a particular objective related to statistics. At the tactical level of statistical thinking, the means of achieving something or for implementing a given strategy is our prime concern. The operational level of statistical thinking involves dealing with statistical data, tools, and analyses.

Putt et al. (2000) presented a hierarchy of statistical thinking slightly different from Dransfield et al. (1999) that was arranged into four levels: idiosyncratic thinking, transitional thinking, quantitative thinking, and analytical thinking. Idiosyncratic thinking would be based on personal judgement and experience. Transitional thinking is somewhat like "reasoning with data" as shown in Bishop and Talbot (2001). Quantitative thinking is sort of quantitative reasoning to form a basis for making judgements or decisions. Analytical thinking is more or less about utilizing given information, formulating problems, and devising a scheme to resolve problems. Pfannkuch (2000) outlined the fundamental elements of statistical thinking that aimed at:

1. understanding data (equivalent to "reasoning about data")

2. transforming statistical data in a meaningful way
3. noticing variation in a real situation that influences statistical strategies to be employed
4. evaluating model accuracy, model validity, model practicality, etc.
5. translating a statistical model into the model of the real context

Prior to any operation for making improvements, a good understanding of the system is a prerequisite. Subsequently, creating a strategy and devising a scheme for making improvements are at the strategic and tactical levels of statistical thinking respectively. Afterwards, thinking about implementation is the next concern in which data are required and statistical tools are used properly. This level of statistical thinking is operational when dealing with data. Obviously, Bishop and Talbot (2001), Mallows (1998), Moore (1997), Putt et al., (2000), and Pfannkuch (2000) tend to focus on tactical and operational levels of thinking; the tasks involved in statistical thinking are broken down into subtasks with sub-goals to serve pedagogical purposes. However, attention should also be given to the strategic level when developing students' statistical thinking because statistical thinking may begin with process rather than data, such as a system may need improvement (Dransfield et al.; 1999; Hoerl & Snee, 2010). As a matter of fact, the approach to statistical thinking adopted can vary for different statistical topics. Hence, it would be difficult to teach generic skills of statistical thinking because the forms of statistical thinking vary somewhat with different approaches.

To develop students' statistical thinking associated with the topic of regression modelling, a teacher in the present study (the author) responded inspiringly because it is a common statistical tool for building models for making predictions. First, a model of statistical thinking in simple regression modelling task was developed in Section II. Second, students' statistical thinking would better be developed through concerted effort in the way making their ideas available via communication to their peers for comment, suggestion, and argument. Their thoughts are thus articulated and statistical thinking would be enhanced. This is grounded in Vygotsky's (1978) socio-cultural theories of learning, emphasizing that language is a tool for communication as well as thinking through which students interact among themselves or with their teacher (Goos, 2009; Li, 2012; Li and Goos, 2017); see Section III for more details. Hence, Section IV would discuss how to foster a learning environment that would stimulate discussions among students to promote thinking. Sections V and VII attempt to address the following two research questions:

1. How well students' statistical thinking was developed?
2. How statistical thinking was developed through collaborative learning?

II. STATISTICAL THINKING IN SIMPLE REGRESSION MODELLING

Although there is no way to guarantee building a good regression model in terms of accuracy and practicality, understanding the interplay of data, statistical thinking, and regression models facilitates progress in data modelling and leads to accurate prediction (see Figure 1). Prior to fitting data to a regression model, it is quite common to study data in terms of its context, measurements and measurement units; statistical thinking follows but back checks patterns, trends, centres, clusters, gaps, outliers, spreads, and variations in data. To commence model building, statistical thinking focuses on classifying the data into independent and dependent variables according to data content. After model building, statistical thinking is about to sort out how to refine the model. Using data to make predictions is our ultimate goal but needing to check model accuracy in terms of the discrepancy of predicted values and given data. The nature of statistical thinking used in regression modelling can be analysed at the strategic, tactical, and operational levels. These three levels provide a useful framework for building and evaluating models.

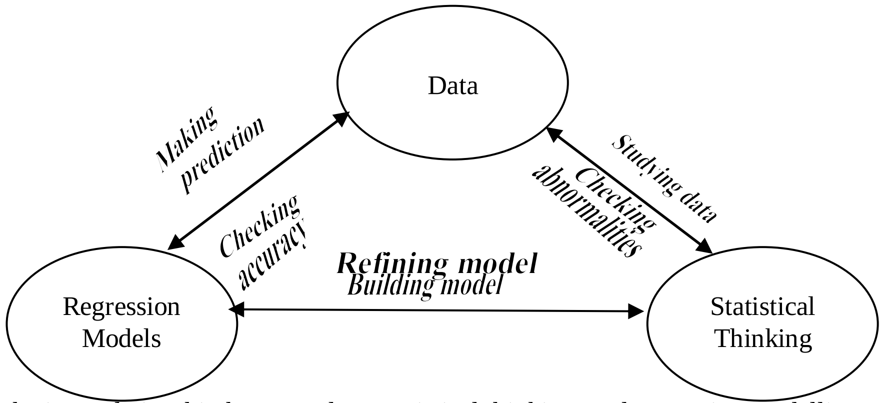

Figure 1. The interrelationship between data, statistical thinking, and regression modelling

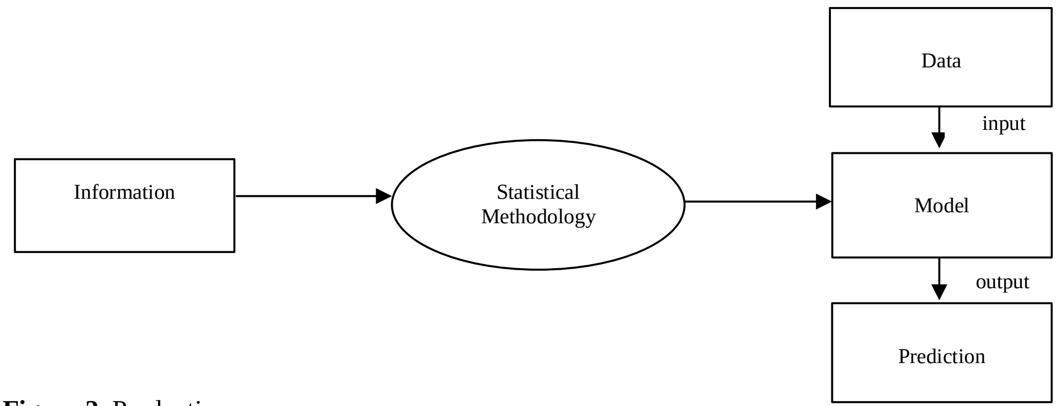

Figure 2. Production process

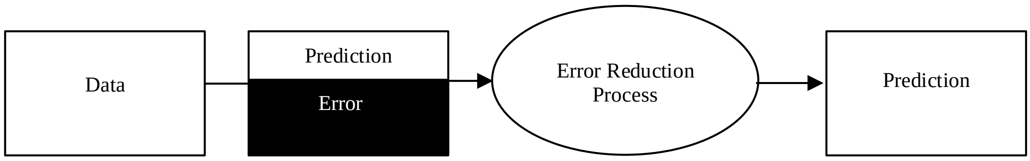

Figure 3. Error reduction

Strategic Level

To make predictions about unknown outcomes based on given data, the prime concern at this stage is to know what is to be predicted and what to do with the data. It is common practice to deduce a model from given data so that predictions can ultimately be made based on the model and the relevant data. This prediction process can be represented by the flow diagram, Figure 2.

When the deduction requires thoughts associated with statistical methodology, this approach is known as statistical modelling. The prediction accuracy relies on the tactics employed in the statistical modelling process.

Tactical Level

In order to make predictions as accurate as possible, it is essential to study the data context and to extract insight from the given data. This enables people to generate ideas to deduce results from the data exploration. The underlying philosophy of making predictions is to turn data into a statistical model that can best describe the relationships within the data. In reality, no model perfectly fits statistical data but possibly incorporates error. The error of prediction is the discrepancy between the observed data and the prediction based on a statistical model. Their relationship can be summarized as a mathematical equation, Error = Data – Prediction and is represented diagrammatically in Figure 3.

In fitting a model to statistical data, error due to inaccurate measurement or random variation of data is inevitable but the error of prediction is expected not to be large. Otherwise, the model is not useful for prediction. For this reason, fitting of data to the model aims at minimizing error (the sum of squared discrepancies between the observed data and the prediction), whereas error reduction is the tactic used in the statistical modelling process (see Figure 3).

Regression modelling is no exception; it seeks to establish a statistical relationship between a dependent variable and one or more independent variables. The simplest way to derive a statistical model from data is to discern visually possible linear trends in the data. This straight-line model is deterministic and may be useful for data prediction. In reality, the data are not exactly located on a straight line. Consequently, fitting a straight line to the data becomes a problem when the data do not follow a linear pattern. In the modelling procedures, the transition from the deterministic model to the stochastic model requires tactics to take care of non-linear patterns and random fluctuation in real-life data; and to decide what steps are involved and in what order the steps should be taken. Regression modelling is an iterative process involving the following tactics:

1. Examine the data to reveal the relationship between variables
2. Select a regression model that can best describe the relationship within the data
3. Fit the proposed regression model
4. Assess how well the regression model fits the data
5. If the data do not fit the model well, model refinement should be undertaken. Model refinement involves synthesizing new ideas with the existing model into an improved model. With these new ideas, go to Step 2 for model refitting (i.e., rebuild the best fitting model for data).

Operational Level

At the operational level, translation of a real-life prediction problem into a statistical problem requires tools and skills for carrying out regression modelling actions. To begin regression modelling, the process of data examination explores the relationship between variables by looking at these three aspects: contextual relations between variables, statistical relations between variables, and functional relations between variables.

The usefulness of a regression model for making meaningful or practical prediction essentially relies on a set of candidate variables suitable for the prediction purposes. Theoretical considerations can be utilized to help in choosing appropriate data for model building. A regression modeler thus requires a good deal of background information on a problem and needs a sound understanding of both the underlying theory and correct selection of data. In selecting such data the modeler judges the usefulness, importance, appropriateness or relevance of data at hand for regression modelling. It is also better to know whether or not the data cover a reasonable and a meaningful range. Further, insight into the data can be gained by mapping between the information in the data and knowledge of problem background. This enables a regression modeler to construct an appropriate regression model in further modelling work.

The statistical models need to be distinguished from deterministic models because statistical models allow for the possibility of error in the description of a relationship. In this way, regression modellers are interested in ascertaining the relationship between dependent and independent variables. One of their concerns is to establish whether or not these two variables are related. If related, they need to look for a data pattern in the graph to see how they are related and how they can be better represented by a mathematical function (for details, see Li & Goos, 2011).

Here the modeller graphs the data points and judges whether the relationship is linear or not. If the relationship turns out to be nonlinear, one must compare the basic pattern to the known generic graphs and attempt to summarize it in a suitable function. It is possible that specific knowledge within a discipline may suggest a certain functional relationship other than linear. In some situations, an appropriate data transformation is employed to transform original data to achieve linearity. A regression analysis is carried out with the transformed linear data, as this is a prerequisite for linear regression modelling. Successful use of data transformation also leads to relatively simple estimation methods, and procedures, and a simpler regression model is ultimately constructed for more practical use.

Using the candidate variables, appropriate models for regression analysis can be suggested and different models would be constructed. In selecting an appropriate regression model, comparisons between different models are made by checking the statistical significance of each regression parameter in each model. The purpose of this check is to determine whether there are any other key variables that could provide predictive power to the model.

When fitting real data to a regression model, it is worth interpreting the meaning of regression parameters to see whether or not their meaning makes sense of reality and matches the underlying rationale. This process may yield one or two potentially useful models that can be used to summarize data in a suitable function and make predictions.

Applying inferential procedures to test the statistical significance of the model is the next concern of a regression modeler. The contribution of each independent variable to the regression model is evaluated, with a statistically significant parameter implying the importance of an independent variable. Apart from the identification of significant regression parameters, interpretation of regression coefficients should be communicated in everyday language commonly understood by lay people. In some cases, the coefficient of an intercept has no practical meaning but only serves as a starting point for prediction. It is also important to assess whether the coefficient of a slope reveals conflict between the empirical model and the underlying theory. The interpretation aims at generating and evaluating arguments about what and what not to believe.

Not surprisingly, any model built on real data may not show an entire representation of reality as such a model takes into account measurement error and individual variability. It is thus necessary to assess how well a regression model fits the data and how well it might be useful for prediction. When more than one tentative regression model is statistically significant and suitable for prediction, the adjusted coefficient of determination, R^2_{adj} (a quantitative measure of how well a regression model fits the observed data) is particularly useful for comparing the regression models. As larger R^2_{adj} corresponds to a regression model that better fits the observed data, the regression model with the largest adjusted coefficient of determination is selected.

Prior to constructing a regression model, four regression assumptions: data linearity, homoscedasticity (constant variance), independence and normality are usually assumed to be valid. However, violation of any of these assumptions may adversely influence prediction results. To detect departures from these four regression assumptions, diagnostic checks on residuals (the difference between the observed and the corresponding predicted values) are made. The selection of the ultimate regression model depends upon the diagnostic results. If the results of the diagnostic checks are unfavourable, remedial actions ought to be taken.

III. LITERATURE REVIEW

Vygotsky's socio-cultural theories of learning (1978) was widely adopted for building social environments in classroom teaching and learning, e.g., Goos (2009), Li (2012), Li and Goos (2015 & 2017), Mogro-Wilson et al. (2015), Negueruela-Azrola and Garcia (2016), as well as Wang et al. (2011). In the study of Mogro-Wilson et al. (2015), it was found that social work students working collaboratively would have access to alternative perspectives on problems at hand through reviewing and interpreting their peers' feedback, thus enhancing understanding or uplifting academic performance. In classroom language teaching, the communication tools like monism, dialectics, cultural mediation, and verbal thinking internalization, which demand human thought would be learnt through social interactions (Negueruela-Azrola & Garcia, 2016). The study conducted by Wang et al. (2011) using socio-cultural theory was to consolidate the talents and experiences of both academics and practitioners in the field of information literacy to generate the basis for undergraduate curriculum development. Socio-cultural theories so far served different academic purpose in these three studies, but none reported the detailed practices of teaching and learning in mathematics classroom.

Nevertheless, the first four studies (Goos, 2009; Li, 2012; Li and Goos, 2015 & 2017) provided more details about socio-cultural theories in a richer context of mathematics and statistics learning and showed the relevance to this present study. Goos (2009) discussed how a community of inquiry comprised of students and a teacher in a secondary mathematics classroom would be established using a socio-cultural framework. The teacher encouraged students to participate in classroom discussions and monitored and moderated the discussions. Through discussions, students communicated their own beliefs, ideas, and understanding, thus making different contributions and generating a more comprehensive view of learning contexts.

The three research articles, Li (2012), Li and Goos (2015) as well as Li and Goos (2017) which drew a closer theme of the present study, conducted quantitative analyses of students' views on social processes of statistics learning within an IT environment. Li (2012) reported that students held positive perceptions and attitudes towards social context of statistics learning, whereas Li and Goos (2015) showed that the students had appreciation of their teacher's supporting and facilitating roles while working collaboratively with groupmates. Both research findings were testified in the study of Li and Goos (2017) by identifying potential factors like teacher's scaffolding assistance and collaborative learning. Unfortunately, none of these articles showed social processes of students' development of statistical thinking that is why the context of the present study is situated.

IV. SOCIAL CONTEXT OF LEARNING

To develop students' statistical thinking, it would be better to adopt a small-group learning approach because Gillies (2021), Mercer (2004), Vygotsky (1978), as well as Zavershneva and van der Veer (2018), who argued that language is a tool for communication and thinking. Students make their ideas available via communication to peers. Language can be used for making thinking explicit so that peers can read and respond. Students' minds would thus be broadened by appreciating others' verbal responses and internalizing these as inner speeches for directing or redirecting task progression or improvement. Hence, the classroom would better be organized to foster an environment facilitating discussion and social interaction.

The students who enrolled in Year 2 of the 3-year Higher Diploma in Applied Statistics and Computing (HDASC) program in a vocational training institute voluntarily participated in this study. The program was to equip students with statistical knowledge and practical skills to solve problems. The HDASC graduates were eligible to join the statistics workforce. Regression modelling was one of the modules offered in the second year of the program; they were taught how: 1) to formulate regression problems; 2) to build regression models; 3) to assess model fitting; 4) to validate or refine the models; and

5) check model practicality.

Apart from fulfilling their prospective job requirements, students' understanding of statistical concepts and methods should be enhanced through the learning of modelling as activating statistical thinking (Stigler & Son, 2018). The topic was taught by the teacher in the present study with an emphasis on social processes of learning because modelling acquires richer and more exhaustive in thinking context through concerted effort and peer communication. The students were divided into small groups in order to provide opportunities to foster teamwork and co-operative working skills that are essential for the statistics workplace; there were twenty-five groups of two students and two groups of three students. Based on their grade point average achieved in their Year 1 Study, a more capable student was grouped with less capable peer(s). These groups characterized asymmetric expertise; similar learning outcomes of each group would be expected. Students within each of these groups naturally sat together or near each other when attending a class held in a lecture theatre and they collaborated on the worksheets in a practice session conducted in a statistical computing laboratory. After class, they were expected to work together, they were asked to complete a regression project on a group basis. Projects enhance students' motivation and engagement with statistics learning because they would see real-life phenomena through a practical application of statistical methods (Teran, 2020).

V. REGRESSION PROJECT

Statistics teaching and learning can be enhanced through project work that would allow students to gain knowledge and skills ready for solving statistical problems in their prospective careers because of the practical nature of statistics (daSilva & Pinto, 2020). The knowledge and skills are associated with the rationale of statistical thinking.

Each group of students was assigned a regression project by the end of the module, "Regression Modelling" so that they would appreciate the relevance and practical use of regression modelling in which they interconnect between statistical concepts. Each group could choose one of these themes of study: traffic and public transport, manpower resources, water consumption, retail business, and import and export trades and downloaded a set of relevant data consisting of a dependent variable (y) and three independent variables (x_1, x_2, x_3) from a computer server.

Students were given a project template in a lecture during which their teacher went through the template together with an elaboration of how to decompose a project into tasks in accordance with a regression modelling workflow and what to do in each task (refer to Tables 2 and 3).

The quality of the projects was evaluated based on the assessment rubrics as shown in Table 2. The rubrics were derived from Biggs and Collis (1982), Bishop and Talbot (2001), as well as Bude (2006) because such evaluation would provide information to both teachers and students how well students reason about data; reason about results; and reason about conclusions so as to improve pedagogy to support learning. The reasoning ability so assessed is associated with the model of statistical thinking developed by Bishop and Talbot (2001). Bude (2006) evaluated students' statistical understanding using three levels: elementary, intermediate and highest achievement. Elementary level evaluates general understanding of statistical definitions and procedures. Intermediate level requires a deeper understanding of statistical data as well as statistical methods. Highest level refers to the skills of justifying and interpreting statistical results. According to Biggs and Collis (1982), students would attain one of the five levels of achievement: "prestructural", "unistructural", "multistructural", "relational" or "extended abstract" (see Table 1). Intuitively, Biggs and Collis (1982) provided exhaustive assessment of students' learning tasks, whereas but Bude's assessment framework (2006) is closely related to the field of statistics education.

Table 1. The assessment framework derived from Biggs and Collis (1982)

Level of achievement	Student performance
Prestructural	Students attempt simple tasks, e.g., selecting an appropriate graphing tool but without utilizing graphic features: titles, labels, scales, axis, and symbols
Unistructural	Students use one relevant aspect, e.g., using one of the graphic features in a statistical graph
Multistructural	Students use several relevant aspects but treat them unrelated or unconnected, e.g., utilizing all the graphic features but treat these as isolated entities and/or unrelated
Relational	Students integrate the relationship between different aspects, e.g., integrating the relationship between the measurement, measurement unit, content and context of data and all the graphic features
Extended abstract	Students deduce relationships such as the qualitative relationship between data

Table 2 presents achievement grades by the nature of regression tasks in accordance with the assessment rubrics in Table 3. Twenty-seven group projects were submitted on time. Among these projects, 74.1% which presented a concise introduction in Task 1 achieved grade A. Approximately 18.5% generally gave a clear introduction but some with imprecise wording or without mentioning variable names of given data, achieved grade B. About 7.4% gave an introduction but lacked quantitative descriptive or specific terms/variable names, thus achieving grade C.

In Task 2, seventeen projects (63.0%) clearly spelt out the objectives of their projects with proper wording, achieving grade A. Among them, some had an imperfection, e.g., using improper quantifiers like mixing up "many" and "much, "volume" and "number", or "sales" and "sales volume". Seven projects (25.9%) achieved grade B showing two similar objectives and inappropriate or imprecise wording, no quantifiers, or unspecific variable names, whereas three projects (11.1%) achieved grade C as terms/variable names were not specified. These first two tasks demanded much thought about how: i) to use given data to formulate a regression problem; ii) to develop the logic of regression workflow; and iii) to set grounds for composing a good written report; statistical thinking at the strategic level was so acquired.

Table 2. Achievement grades by the nature of regression tasks

Task	Achievement grade				
	A	B	C	D	E
1. To give an introduction	20	5	2	0	0
2. To formulate research objectives	17	7	3	0	0
3. To check the reliability of given data	14	10	3	0	0
4. To build simple linear regression models	21	5	1	0	0
5. To justify the model significance	22	4	0	1	0
6. To validate the regression model	20	7	0	0	0
7. To summarize regression results and draw a conclusion	18	3	5	1	0
8. To give an verbal presentation	6	8	13	0	0

Statistical thinking in Tasks 3, 4, 5, and 6 at the tactical and operational levels focuses on data analysis. Scholars like Bargagliotti et al. (2020, p.3) stated "data serve models of reality", so the third task is about valuing appropriate and reliable data that is a prerequisite for substantiating inferential statistics (Schum, 2001). All the projects, 100% correctly checked the context, measurements, and measurement units of given data and evaluated the linearity in each of these three pairs of data, namely y (dependent variable) and x_1, y and x_2, and y and x_3. In addition, all the projects selected independent variable(s) among x_1, x_2, and x_3 that would exhibit stronger linear relationship(s) with y by means of scatterplots and/or correlation coefficients to confirm a linear relationship between y and among x_1, x_2, and/or x_3.

Fourteen projects (51.9%) showed sound judgement of data, thus achieving grade A when attempting Task 3. Few of them displayed exceptionally good work like justifying outliers, solving outlying problems successfully, and discussing the cause(s) or consequence(s) of an outlier. When students had doubts about a linear relationship between a dependent and an independent variable owing to a weak correlation pattern on a scatterplot, they adopted a quantitative approach, i.e., either using a correlation coefficient or conducting a hypothesis testing. Ten projects (37.0%) achieved B owing to a minor fault. First, few projects missed reporting one or two parts of data examination, like neither speculating potential outlying observation(s) nor non-symmetric data. Second, few projects had an incorrect speculation about outlying observation(s) or overstated a controversial outlier. Three projects (11.1%) achieved C owing to two or more flaws, such as overlooking an outlier or skewed data, not fixing outlying problem(s), not quantifying correlations; not linking claims with results, and not discussing statistical findings clearly.

Model construction in Task 4 executes under these four regression assumptions: linearity, homoscedasticity, independence, and normality of data. Statistical interaction, x_1x_2, x_1x_3, x_2x_3, or $x_1x_2x_3$ were not explicitly modelled in the simple regression context because the notion of these interactions had not been taught but would be discussed later in the sophisticated topic of multiple regression modelling. Among the models students had built, they chose the model with the largest R^2 or R^2_{adj} in a simple linear regression model. Twenty-one projects (77.8%) achieved grade A by utilizing both graphical and computational tools to establish sound evidence in approving or disapproving of potential model(s). Among these projects, some gave model implications or interpretations and very few improved model fitting by removing outliers one by one. Five projects (18.5%) achieved grade B because some justifications were somewhat ambiguous like lacking axis labels, missing significant statistical results, or providing inconsistent arguments. Only one project (3.7%) did poorly and achieved grade C owing to a wrong justification of model construction.

Table 3. Project assessment rubrics

Tasks	Performance descriptors, i.e., the extent to which a project has shown	Assessment grade
1, 2, 7	Very good writing with very clear, logical, and persuasive expressions	A
	Good writing with quite clear, logical, and persuasive writing expressions	B
	Satisfactory writing with fairly clear, logical, and persuasive expressions	C
	Poor writing with some unclear, illogical or non-persuasive expressions	D
	Very poor writing with unclear, illogical, and non-persuasive expressions	E
3	Comprehensive examination of data using proper statistical tools with reasoned judgement	A
	Comprehensive examination of data using proper statistical tools incorporating judgement with some grounds	B
	Comprehensive examination of data using proper statistical tools incorporating judgement with a few grounds	C
	Comprehensive examination of data using improper statistical tools without reasoned judgement	D
	Comprehensive examination of data using improper statistical tools without judgement	E
4, 5	Comprehensive knowledge of statistical tools and Excel programming	A
	Sound knowledge of statistical tools and Excel programming	B
	Adequate knowledge of statistical tools and Excel programming	C
	Basic knowledge of statistical tools and Excel programming	D
	Scarce knowledge of statistical tools and Excel programming	E
5, 6	Complete justifications or validation	A
	Partial justifications or validation	B
	Weak justifications or validation	C
	Unsounded/wrong justifications or validation	D
	Irrelevant justifications or validation	E
8	Very good verbal presentation	A
	Good verbal presentation	B
	Adequate verbal presentation	C
	Poor verbal presentation	D
	Very poor verbal presentation	E

Both Tasks 5 and 6 required students to think thoroughly by checking the statistical significance of regression estimates, β_0 and β_1 and the model validity. In Task 5, Twenty-two projects (81.5%), which correctly evaluated the inferential force of statistical evidence and utilized the evidence to defend the model(s) they had built or substantiated an allegation against the model(s) achieved grade A. Four projects (14.8%) achieved grade B because some statistical arguments were unconvincing or controversial owing to marginally significant regression estimates. In addition, a few projects showed regression estimates that were insignificant but did not find out the cause(s) of a poor model fitting, over interpreted or did not report an insignificant regression estimate. Only one project (3.7%) did poorly and achieved grade D because a critical mistake arose from incorrect evaluation of the inferential force of statistical evidence.

By validating the model in Task 6, they constructed arguments in favor or not of the four regression assumptions using regression diagnostics tools, i.e., residual plots and normal probability plots. Twenty projects (74.1%) correctly used the tools and successfully decoded data patterns on the plots to confirm or disconfirm the assumptions. Besides, few projects identified certain troubles arising from a departure or a mild departure of one of the assumptions and suggested how to fix an invalid regression assumption. Seven projects (25.9%) achieved grade B because unconvincing justification of data linearity was provided, students could not decode residual patterns, or failed to identify an outlier. These validation results form the basis for an optimal model or offer clues in a model refinement process. Few projects failed to defend the linearity assumption because the linear pattern of data was distorted due to data scales.

Statistical thinking beyond Excel programming, Task 7 is about to organize and connect individual regression results to compile a comprehensive report in which regression results need to be translated into physical or practical meanings in our everyday language. Eighteen projects (66.7%) which achieved grade A gave a complete and valid report with credible lines of reasoning about the model building and links among inferential tools, inferential results, and statistical evidence irrespective of whether or not the model they had built was reliable. They made use of R^2 or R^2_{adj} to grade the extent to which the model would fit the given data, together with a recommendation or drawbacks of using the model for prediction. The drawbacks are about a model being unable to make accurate or reliable predictions, a linear regression approach being unable to fit seasonal or nonlinear data, or an outranged $\hat{\beta}_0$ in a simple regression model being unable to provide a meaningful or realistic context. In case of a poor model fitting, they suggested how to refine or revise the model like replacing given data with the data more relevant to their research objectives or appending more data. They further illustrated the practical significance and usage of the model and interpreted the meaning of $\hat{\beta}_0$ and $\hat{\beta}_1$ relating to the context and measurement units of data. A project also showed complete and consistent arguments, achieving grade A although a technical flaw arose from the third task. Three projects (11.1%) achieved grade B as not fulfilling some assessment criteria. Few projects with interpretations did not link the context and measurement units of data, showed a wrong data label in an interpretation of $\hat{\beta}_0$, presented models incorporating an insignificant $\hat{\beta}_0$, or did not interpret $\hat{\beta}_0$ and $\hat{\beta}_1$. Five projects (18.5%) achieved grade C because few projects under or over reported the significance of a model or regression parameter(s). In addition, only one project (3.7%) which showed an argument inconsistent with a regression result achieved grade D. The lexical content in the written reports showed logical flow, thinking, and reasoning although some imperfections or mistakes arose from not having good command of English language because the students were not native English speakers.

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 instructional scaffolding.

All groups of students finished a verbal presentation of their projects within the time allowance in front of class in a computing laboratory. Almost all groups of students made a clear and smooth presentation because they were well prepared. Before presentations, students within each group decided which of the tasks in Table 2 they would present. They put their presentation materials including tables and graphs in PowerPoint slides in an orderly way. During presentations, the labor of division in each group was evident from co-operative and coordinated roles in carrying out presentation tasks. When one student was presenting a project, his/her groupmate(s) commanded the PowerPoint to move synchronously with the presentation. Non-verbal interactions within groups displayed in the circumstances where a student asked his/her groupmate(s) to move PowerPoint slides forward or backward or highlight certain parts of a table or graph using a mouse pointer. Some students had sound skills of question handling when being asked by their teacher; they gave quick responses and convincing arguments. The questions were to offer presenters directives or an opportunity to elaborate, clarify, or defend reasoned arguments. If students could not answer the questions correctly and completely, groupmate(s) supplemented their answers or gave hints or the groups had a quick and quiet discussion. Verbal interaction that arose from these situations was positive and conducive to peer learning. Both verbal and written presentations of their projects showed that each group of students worked cooperatively after class and peer collaboration presented an ideal context for promoting the development of statistical thinking associated with regression modelling.

The quality of verbal presentation skills was assessed based on presentation flow, speech delivery, question handling, time management, and effective use of tables and graphs. Each HDASC student played a significant role and gave nearly equal contributions in the project work. Among the twenty-seven projects, six (22.2%), eight (29.6%), and thirteen (48.2%) did very sound presentation (grade A), sound presentation (grade B), and adequate presentation (grade C) respectively.

Projects attained different grades in different regression tasks (see Table 2). The grading of each task was used to compile an overall grade to summarize group performance of each project. Among the twenty-seven projects, fifteen (55.6%) achieved an overall grade close to A, whereas twelve (44.4%) achieved a grade slightly higher than B. Project achievement is probably related to a productive collaboration during which students contributed towards regression tasks and monitored and revised regression solutions closely.

VI. VALIDITY OF PROJECT ASSESSMENT

The External Examiner of this HDASC program was Professor of Statistics and the Program Leader of B.Sc. in Applied Statistics and Computing program at a university in Hong Kong, China. He was very familiar with the ability and learning attitudes of students in Hong Kong where the present study is situated. He checked assessment instruments aligned with the pre-set learning objectives, and learning outcomes of the HDASC students. He had taken a sample of the projects after academic assessment for checking to ensure the projects were assessed fairly and consistently. The only concern he raised was about how much each individual student had contributed to his/her own group project in terms of marks. Each student was assessed during project supervision as well as in their oral presentation. In addition, peer assessment was made by each student to inform the teacher about how much his/her groupmate(s) had done in group project. Each group project was assessed in terms of its quality by assigning a grade, so there is no large variation in grades among students within a group of two or three. In addition, a small percentage of project grades was counted towards each student's overall achievement, thus not affecting the final grade significantly.

VII. STUDENT FEEDBACK

Section V had addressed the first research question of how well students' statistical thinking was developed. The next concern was how statistical thinking was developed through collaborative learning because peer collaboration does not arise by just putting students in small-group learning situations. In fact, there is a need to foster an environment that would allow students to engage with peers (Nussbaum et al., 2009). A questionnaire-based survey was therefore conducted to solicit feedback from a whole class of 56 HDASC students when working in groups, questions were asked to appraise peer support, assistance, and performance. Some questions provided three response categories: positive, neutral, and negative. A few questions were open-ended, so the students could elaborate on why they gave positive, neutral, or negative opinions through answering these questions.

To ensure the survey data accurate, two phases of data validation were carried out. The first phase of data validation took place during data processing. The second phase of data validation was to check whether there were missing data, meaningless data range, data inconsistency, and undefined data codes. Statistical analysis of the survey data was performed by means of descriptive statistics and statistical tables, graphs, and charts. Students' responses to open-ended questions were generally extensive and supplementary to why they chose a specific response category as their answer. The responses which were similar were grouped so that the reasons could be further summarized.

Among 56 students, 51.8% reported having much interaction, 44.6% gave a neutral response, and 3.6% reported having little interaction. Forty-seven students (83.9%) found learning partnership harmonious, nine students (16.1%) gave a neutral response but none (0%) gave a negative response. Forty-one students (73.2%) said their groupmate(s) made learning fun while fifteen students (26.8%) gave a neutral response and none (0%) gave a negative response. Forty-five students (80.4%) had group harmony but none did not, eleven students (19.6%) gave neutral responses. They described the above positive feelings of learning with groupmate(s) in a number of common ways, mostly related to their experience of learning in a pleasant atmosphere with mutual support. Some of the support was socially oriented; they listened to each other and expressed concern or provided counselling. The support could also be in the cognitive domain like knowledge sharing, intellectual exchanges, and so on.

Most students tentatively ventured the opinion that the project would be difficult to complete but fifty-one students (91.1%) found that their groupmate(s) co-learning with them rather than competing and six students (8.9%) gave a neutral response. They set and achieved a common goal of learning; they collaborated on laboratory worksheets, assignments, a project, and so forth. In the context of co-learning, they shared knowledge in the domain of statistical logic, statistical thinking, and statistical reasoning with their groupmate(s). They also shared knowledge around communication skills associated with interpretation of statistical results, enhancement of statistical reports in a written or a verbal form in addition to procedural and technical knowledge i.e., operational level of statistical thinking. The discussions they held was a form of productive interaction where a student evaluated ideas his/her groupmate(s) brought forward in order to come up with a joint decision leading to an answer. Yet, if they did not reach a consensus as a basis for joint progress, they sought the teacher's assistance. Through teacher scaffolding of the process of inquiry towards hypothesizing, interpretation, and reasoning, students gained an appreciation of their teacher as being an active and communicative participant in learning and their subsequent talk led to negotiation with groupmate(s) about how to use statistical evidence to substantiate their reasoned arguments. Students within groups claimed to have much interaction with groupmate(s) and found their peer relationships were harmonious as well as contributing to a positive learning atmosphere, thus demonstrating group cohesion.

Fifty-three students (91.4%) believed that communications with groupmate(s) were conducive to their learning process with the following reasons. Through communication, they determined goals; devised and regulated strategies for problem solving; stimulated one another to think; clarified misconceptions or misunderstanding; and resolved conflicting views. The learning progress was characterized by a high degree of mutual involvement and collaborative interaction. Only one student (1.7%) had a negative opinion because a less capable groupmate did not assist their learning, and four students (6.9%) gave a neutral response. However, it is worth noting that there is a close link between social interaction and development of statistical thinking.

VIII. CONCLUSION

Project achievement is probably related to group processes as students within each group directed their own projects and regulated their strategies with concerted effort and group decisions. First, this is evident from the quality of their verbal and written presentations. Second, the teacher provided supervision to students for developing skills of recognizing and solving problems and found that there were a series of interpersonal and communication skills used by the students within each group when collaborating in formulating research objectives; exploiting and regulating statistical methodology; as well as report writing.

Third, group processes are evident from the survey findings and are characterized by positive working and interpersonal relationships, and productive interaction when doing projects collaboratively irrespective of whether homogeneous or heterogeneous grouping of students. A few groups are more or less homogeneous with students of equal expertise, they provided each other with learning assistance and support. Most project groups were heterogeneous; a more capable student provided scaffolding assistance to less capable peer(s). Students within a group utilized proper statistical tools, discussed regression heuristics, evaluated the inferential force of evidence, constructed reasoned arguments, and complied a written report. Productive collaboration would arise from exchanges of personal experiences, ideas and/or insights through verbal and/or non-verbal interactions. Verbal interactions can be affective, cognitive, or regulative in nature, whereas non-verbal interactions like gestures may intend or may not intend to convey a message. The development of statistical thinking and reasoning is evident from social interaction among students that is characterized by a collaborative interaction among peers in the way that the students talked about how to address a question of common concern based on their own understandings, opinions, judgments or perspectives through verbal exchanges. These probable findings are evident from a series of observation studies conducted in the same classroom by the author of this paper and his colleague (see Li, 2015; Li, 2021; Li & Goos, 2018).

The major implication of this study is thus that statistics learning tasks would better be contextualized in authentic activities through which students derive understandings and interpretations of learning tasks. In addition, these tasks should be designed to engage students in collaborative learning with their active involvement and participation.

ACKNOWLEDGEMENT

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

References

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre K-12 guidelines for assessment and instruction in statistics education II-A framework for statistics and data science education*. American Statistical Association and National Council of Teachers of Mathematics. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf.

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press. https://doi.org/10.1016/C2013-0-10375-3

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). International Association for Statistical Education and International Statistical Institute.

Bude, L. (2006) Assessing students' understanding of statistics. In A. Rossman & B. Chance (Eds.), Proceedings of the 7th International Conference on Teaching Statistics. International Association for Statistical Education. https://www.stat.auckland.ac.nz/~iase/publications/17/6G3_BUDE.pdf

Chance, B. (2002). Components of statistical thinking and implications for instruction and assessment. *Journal of Statistics Education, 10*(3), 1-14. https://doi.org/10.1080/10691898.2002.11910677

daSilva, M. P. M., & Pinto, S. (2014). Teaching statistics through learning projects. *Statistics Education Research Journal, 13*(2), 177-186. https://doi.org/10.52041/serj.v13i2.289

Dransfield, S. B., Fisher, N. I., & Vogel, N. J. (1999). Using statistics and statistical thinking to improve organization performance. *International Statistical Review, 67*(2), 99-121. https://doi.org/10.2307/1403389

Hoerl, R. W., & Snee, R. (2010). Statistical thinking and methods in quality improvement: A look to the future. *Quality Engineering, 22*(3), 119-129. https://doi.org/10.1080/08982112.2010.481485

Frischemeier, D., Kazak, S., Leavy, A., Meletiou-Mavrotheris, M., & Paparistodemou, E. (2022). International perspectives of early statistical thinking: Comparison of primary school curricula in different countries. In S. A. Peters, L. Zapata-Cardona, F. Bonafini & A. Fan (Eds.), *Proceedings of the 11th International Conference on Teaching Statistics*. International Association for Statistical Education. https://iase-web.org/icots/11/proceedings/pdfs/ICOTS11_123_FRISCHEM.pdf

Gillies, R. M. (2012). Promoting reasoning, problem-solving and argumentation during small group discussions. In R.M. Gillies (Ed.), *Pedagogy: New developments in the learning sciences* (pp. 131-150). Nova Science Publishers.

Li, K. W. (2015). Analysis of discourse practices in statistical computing laboratory. In S. Gerland (Ed.), *Proceedings of the Second 21st Century Academic Forum Conference*, (pp. 249-263). Harvard University, USA. www.21caf.org/uploads/1/3/5/2/13527682/20_hrd-531-li_ed2_fmt_logo.pdf

Li, K. W. (2021). Developing regression heuristics through collective argumentation in computing laboratory. In R. Helenius (Ed.), *Proceedings of the International Association for Statistical Association 2021 Satellite Conference*. International Association for Statistical Education. https://iase-web.org/documents/papers/sat2021/IASE2021%20Satellite%20124_LI.pdf

Li, K., & Goos, M. (2011). Reinforcing students' correlation comprehension. *The International Journal of Learning, 17*(11), 261-274.

Li, K. W., & Goos, M. (2018). An analysis of patterns of classroom talk within an IT environment. In M. A. Sorto, A. White & L. Guyot (Eds.), *Proceedings of the 10th International Conference on Teaching Statistics*. International Association for Statistical Education. http://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_3G1.pdf

Mallows, C. (1998). The zeroth problem. *The American Statistician, 52*(1), 1-9. https://doi.org/10.2307/2685557

Mercer, N. (2004). Sociocultural discourse analysis: Analyzing classroom talk as a social mode thinking. *Journal of Applied Linguistics and Professional Practice, 1*(2), 137-168. https://doi.org/10.1558/japl.v1.i2.137

Mogro-Wilson, C., Reeves, M. G., & Charter, M. L. (2015). Applying sociocultural theory to teaching doctoral social work students. *Journal of Teaching in Social Work, 35*(1-2), 151-165. https://doi.org/10.1080/08841233.2014.976354

Moore, D. S. (1997). New pedagogy and new content: The case of statistics (with discussion). *International Statistical Review, 65*(2), 123-137. https://doi.org/10.2307/1403333

Negueruela-Azrola, E., & Garcia, P. N. (2016). Sociocultural theory and the language classroom. In G. Hall (Ed.), *The routledge handbook of english language teaching* (pp. 295-309). Routledge.

Nussbaum, M., Alvarez, C., McFarlane, A., Gomez, F., Claro, S., & Radovic, D. (2009). Technology as small group face-to-face collaborative scaffolding. *Computers and Education, 52*(1), 147-153. https://doi.org/10.1016/j.compedu.2008.07.005

Pfannkuch, M. (2000). A model for the promotion of statistical thinking? In J. Bana & A. Chapman (Eds.), *Proceedings of the Twenty-third Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 503-510). Mathematics Education Research Group of Australasia.

Putt, I., Perry, B., Jones, G., Thornton, C., Langrall, C., & Mooney, E. (2000). Primary school students' statistical thinking: A comparison of two Australian States. In J. Bana & A. Chapman (Eds.), *Proceedings of the Twenty-third Annual Conference of the Mathematics Education Research Group of Australasia* (pp. 519-526). Mathematics Education Research Group of Australasia.

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

Snee, R. D. (1990). Statistical thinking and its contribution to total quality. *The American Statistician, 44*(2), 116-121. https://doi.org/10.2307/2684144

Stigler, J. W. & Son, J. Y. (2018). Modelling first: Approach to teaching introductory statistics. In M. A. Sorto, A. White & L. Guyot (Eds.), *Proceedings of the 10th International Conference on Teaching Statistics*. The International Statistical

Institute. https://icots.info/10/proceedings/pdfs/ICOTS10_9C3.pdf?1531364299

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. https://doi.org/10.52041/serj.v8i2.394

Teran, T. E. (2020). Advantages of applying the project method to teaching of statistics. In P. Arnold (Ed.), *The Proceedings of 2020 IASE Roundtable Conference*. International Association for Statistical Education. http://www.iase-web.org/documents/papers/rt2020/IASE2020%20Roundtable%2036_TERAN.pdf

Vygotsky, L. S., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. 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., & Neal, D. (2012). Preparing students for decision-making in the 21st 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. 89-114). Mathematics Education Research Group of Australasia.

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