It's a good score! Just looks low: Using data-driven argumentation to engage students in reasoning about and modelling variability

Orlando González, Somchai Chitmun

It's a good score! Just looks low: Using data-driven argumentation to engage students in reasoning about and modelling variability

Orlando González [1] and Somchai Chitmun[2]

[1]Assumption University, Graduate School of Human Sciences, Bangkok, Thailand; ogonzalez@au.edu

[2]Tepleela School, Bangkok, Thailand; schitmun@hotmail.com

This paper describes senior high school students' different ways of reasoning about and modelling variability, drawn from the initial phase of an ongoing study. A group of 26 Thai Grade 12 students was engaged in a statistical investigation by addressing a socially open-ended problem based on an unsorted set of varying test scores. Through engagement with this problem, students were able to experience data-driven argumentation and data modelling. From analyzing the answers provided by participants, four different categories of reasoning about variability (i.e., value-relation, magnitude-comparative, proportional, and distributional reasoning) and 11 ways of modelling variability in the context of the given situation were identified. The study also revealed that participants are somehow good proportional reasoners in general, but weak distributional ones.

Keywords: Argumentation, socially open-ended problems, reasoning about variability, variability modelling, statistical investigations.

Introduction

Good citizenship requires being statistically literate. Statistical literacy includes making sense of real-world data, the use of evidence-based arguments and critically evaluating data-based claims (Garfield & Ben-Zvi, 2008). Being critical is essential as claims and arguments are usually presenting selective information to convince another person to adopt or reject a specific point of view (Gal, 2004). On this regard, it has been reported in the literature that data-driven argumentation—which requires using, revising, manipulating, structuring and representing data to explain or persuade others—is very important for data modelling and informal inference (Garfield & Ben-Zvi, 2008; Shaughnessy, Ciancetta, & Canada, 2004).

Such importance has been acknowledged by recent reforms to the mathematics curriculum in many countries, Thailand being an example. In fact, according to the current Thai Basic Education Core Curriculum (MOE, 2008), Grade 12 graduates must be able, among other things, to analyze data and apply the results to express views and persuasive arguments using accurate and appropriate language. However, teacher's guides do not include specific instructional suggestions on this matter.

Under this scenario, socially open-ended problems (Shimada & Baba, 2015) emerge as an appealing and plausible way of providing students with the possibility of using, handling and interpreting data to inform argumentation and decision-making. This kind of problem has been reported as an instructional way to challenge students to structure variability among repeated observations of the same event, to model variability, and to engage in data-driven argumentation (González & Chitmun, 2017). The purpose of the present paper is to answer the following research question: what are the ways, if any, by which Thai senior high school students reason about and model variability, when they are asked to provide persuasive explanations and arguments based on data analysis, in the context of a socially open-ended problem? A lesson implementation of a socially open-ended problem will be discussed and analyzed, and some conclusions will be given based on its results.

Practicing data-driven argumentation as advocacy

Argumentation refers to discourse for persuasion, logical proof, and evidence-based belief, and more generally, discussion, in which there is reasoning on data, doubt about a standpoint or a disagreement with it, and frequently a question about whether, or at least why, a standpoint is worthy of acceptance (Blair, 2003; Garfield & Ben-Zvi, 2008). Argumentation can take two forms: inquiry and advocacy (Toulmin, Rieke, & Janik, 1984, as cited in Watson, 2018). The practice of argumentation as inquiry presupposes the act or process of deriving conclusions from data, whereas the practice of argumentation as advocacy presupposes the questioning of a standpoint, objections to a person's arguments, and arguments against the standpoint a person is supporting (Blair, 2003). Involvement with statistics in daily life is often about judging claims and arguments of others who practice statistics to some degree or another (Gal, 2004; Watson, 2018). Therefore, arguing against claims made by others (e.g., "your mathematics test score was awful!") is likely to require data-driven argumentation associated with advocacy. This involves the ability to advocate against the claim at hand with some evidence grounded in data, in order to make an alternate claim (Watson, 2018).

Statistical investigations and data-driven argumentation

A statistical investigation is a process comprised of the following five phases: Problem, Plan, Data, Analysis, and Conclusion (PPDAC), regarded as one of the four central dimensions of statistical thinking used in statistical inquiry of real-life problems (Makar & Fielding-Wells, 2011). By engaging in statistical investigations, students are able to fully experience statistical processes such as acknowledging, modelling of and reasoning about variability, data representation, data reduction, data-driven argumentation, decision-making, and informal inference (Watson, 2018).

In order to appropriately conduct statistical investigations in the mathematics classroom, teachers and students need to use problems worthy of investigation, which involve statistical questions, a real-life context, and the following traits: (1) interesting, challenging, and relevant; (2) statistical in nature; and (3) ill-structured and ambiguous (Makar & Fielding-Wells, 2011). Addressing questions with these characteristics demands critical thinking skills and the practice of argumentation as advocacy, because questions developed for statistical investigations usually challenge and question, ambiguously or without proper statistical foundation, the standpoint a person is supporting (Blair, 2003; Gal, 2004; Watson, 2018). Thus, undergoing a statistical investigation requires the ability to challenge statements made in the given context and advocate, with evidence, for an alternate claim (Watson, 2018); in other words, requires the practice of data-driven argumentation as advocacy.

Socially open-ended problems as a trigger for data-driven argumentation

Socially open-ended problems (Shimada & Baba, 2015) are problems embedded in a real-life context, familiar to the students and, by extending the traditional open-ended approach (Becker & Shimada, 1997), have been developed to elicit and address students' mathematical values (e.g., visual appeal, parsimony, abstraction, systematic reasoning), social values (e.g., compliance with the law, fairness, compassion, equity), and personal values (e.g., persistence, integrity, friendliness) through modelling and argumentation. So, from the past section, it is possible to identify clear similarities between socially open-ended problems and the problems developed for statistical investigations.

In order to ensure triggering data-driven argumentation through the use of socially open-ended problems, the real-life context in which the problem is embedded must be statistical in nature (i.e., must enable students to collect, represent, reduce and/or interpret data to address a statistical question and reach reasonable conclusions under uncertainty; González & Chitmun, 2017). Thus, arguments from addressing the posed problem will have the four building blocks to make a good argument (Blair, 2003; Garfield & Ben-Zvi, 2008, p. 275): (1) A clear claim (and a counterclaim) we are making and/or anticipating; (2) data to support our argument; (3) evidence that the data are accurate and reliable, and (4) a good line of reasoning connecting the data to our argument.

The purpose of this study was not only to engage students in data-driven argumentation, but also to challenge them to reason about and model variability. To that end, the problem should provide students with opportunities to display, structure and model variability among observations of the same event, to make accessible and meaningful to students characteristics of distribution (e.g., center, shape and spread; Lehrer & Schauble, 2004; Mulligan & English, 2014; Petrosino et al., 2003).

Research methodology overview

Participants

On August 31, 2018, data were collected from a Grade 12 mathematics class (17- to 18-year-old youths) in a large public high school in Bangkok. The second author was the classroom teacher. A total of 26 students (boys=3, girls=23) were administered the task shown in Figure 1.

Protocol and data gathering procedure

The task chosen for this study was an adaptation of the "Let's think of an excuse for the test score" problem (Tamaoki, 2014, p. 120, see Figure 1). This task, a socially open-ended problem, requires from students to put in practice data-driven argumentation as advocacy, because they have to consider the variability in the given empirical distribution of scores and then generate additional arguments to explain Malee's mother why her test score is not necessary such a bad result in the given context.

LET'S THINK OF AN EXCUSE FOR THE TEST SCORE

The numbers below are the results achieved by a certain class on a mathematics test (100 points maximum), sorted by student number in the class attendance sheet. Malee is the only one who got 33 points. Only by what this score tells, she will be scolded by her mother. What could Malee possibly say to avoid a scolding for her mathematics test score?

25	28	45	44	41	28	58	88	100	21	28
16	50	50	45	**33**	21	22	24	25	26	28
30	45	28	23	25	22	77	100	26	58	26
14	12	69	28	18	53	100				

Figure 1: Task used in this study (adapted from Tamaoki, 2014)

This task was collected from a Japanese junior high school teacher during a previous study conducted by the authors (González & Chitmun, 2015). In order that Thai students were able to respond to this task without being distracted by names or contexts unusual or non-existent for them, two changes regarding the general cultural context were made when translating it into Thai: a Thai name for the student and “her mother” (as opposed to “her family” in the original task) were chosen for this version. The adapted task was then printed on single paper sheets and distributed among the students, who individually engaged in solving it after being instructed by the teacher. Fifty minutes were given to complete the task. Students were not assisted in any way that could influence their responses.

Analysis of the collected data

Both authors used open coding and descriptive coding to analyze the information provided by the students, and to group similarly-structured answers into categories for analysis. After consensus between the authors, this process led to four categories on students’ reasoning about variability in the given task. Also, students’ ways of modeling variability were categorized. The authors also agreed on the placement of all the 26 responses into the categories previously created.

Empirical findings and discussion

Participants’ types of reasoning about variability

Students’ reasoning about variability fell into four coding categories: value-relation, magnitude-comparative, proportional, or distributional reasoning. Table 1 shows the percentages of students under each of these categories. Students were allowed to provide more than one argument to present Malee’s low test score as not such a bad result, resulting in a total percentage of more than 100%.

Category of reasoning	Description	%
Value-relation	Arguments using only comparative or ordinal phrases to indicate the difference between Malee’s score and particular data values within the data set.	46.2
Magnitude-comparative	Arguments using both comparative phrases (e.g., “greater than”, “less than”, “more”, “less”) and quantified expressions (e.g., “7 points”) to indicate the difference between Malee’s score and particular data values within the data set.	7.7
Proportional	Arguments using, implicitly or explicitly, the ratio, the actual percentage or the proportion of elements having a particular characteristic within the distribution of outcomes.	100
Distributional	Arguments combining two or more key features of the distribution of outcomes (e.g., centers, shape, skewness, density, outliers, and spread/variability).	15.4

Table 1: Percentage of students using each coding category in this study

Students whose responses included arguments using only comparative (e.g., “greater than”, “less than”, “more”, “less”) or ordinal (e.g., “third”, “fourth from”) phrases to indicate the difference or position between Malee’s score and either some center (i.e., the mean or mode) or a particular data value serving as reference point (e.g., 100 points), fell into the *value-relation reasoning* category. A total of 12 students (46.2%) fell into this category. The next quotes illustrate this category.

Student 20: The mode was 28 points. Malee’s test score is higher than the mode.

Student 26: I got 33 points, which is the 6^{th} best score among all the students who failed the test.

Students whose responses included arguments using both comparative phrases (e.g., "greater than", "less than", "more", "less") and quantified expressions (e.g., "7 points") to indicate the difference between Malee's score and either some measure of center (i.e., the mean or mode) or a particular data value serving as reference point (e.g., 50 points), fell into the *magnitude-comparative reasoning* category. Only 2 (7.7%) out of 26 students fell into this category. Examples of this category follow:

Student 14: Malee got 33 points, which is lower than the passing criterion by 17 points.

Student 15: Tell her mother that Malee's score is lower than the mean only by 7 points.

Students whose responses included arguments using, implicitly or explicitly, the ratio, the actual percentage or the proportion of elements having a particular characteristic within the distribution of outcomes (e.g., the percentage of students who passed and/or failed the test), fell into the *proportional reasoning* category. All 26 students (100%) gave answers falling into this category. The following two sub-categories were identified under this category: *implicit proportional reasoning* (i.e., responses including arguments suggesting an implicit consideration of sample proportions, population proportions, or percentages); and *explicit proportional reasoning* (i.e., responses including arguments explicitly mentioning the ratio, the actual percentage or the proportion of elements having a particular characteristic within the distribution of outcomes).

In this study, a total of 11 students (42.3%) were categorized as implicit proportional reasoners. The next quote illustrates the category "Implicit proportional reasoning."

Student 1: Tell mom that Malee's score is more than the median, and there were only 16 students who got more than 33 points and 23 students who got less than 33 points. Therefore, her score is somehow satisfying under these criteria.

In this study, a total of 15 students (57.7%) were categorized as explicit proportional reasoners. The next quote illustrates the category "Explicit proportional reasoning."

Student 3: From the box-plot diagram, it was found that Malee's score (33 points) is higher than … the scores of 50% of the total students but not reach the score of the 75%.

Students' responses coded as *distributional reasoning* included arguments integrating two or more key features of the distribution of scores, such as centers (e.g., mean, modes, proportions), shape (e.g., skewness), spread (e.g., range, standard deviation) and outliers. A total of 4 students (15.4%) fell into this category. The next quotes illustrate the category "Distributional reasoning."

Student 6: From interpreting the standard deviation, Malee's score lies to the left of the mean, where most of students' scores are. However, Malee's score is in the range between Q_2 and Q_3, which is higher than 50% of students who got similar scores.

Student 12: Most of the students' scores skew to the left of the median.

Participants' ways of modelling variability

Eleven ways of modelling variability by the students were identified (see Table 2). Students were allowed to represent the data in more than one way, resulting in a total percentage of more than 100%.

Model code	Type of model	Model description	Frequency (%)
M_1	Visual	Sorted data array	13 (50)
M_2	Visual	Stem-and-leaf plot	7 (26.9)
M_3	Visual	Boxplot*	7 (26.9)
M_4	Visual	Pie chart	1 (3.8)
M_5	Visual	Bar graph	1 (3.8)
M_6	Visual	Visually estimated skewness	2 (7.7)
M_7	Mathematical	Mean	4 (15.4)
M_8	Mathematical	Median	8 (30.8)
M_9	Mathematical	Mode	8 (30.8)
M_{10}	Mathematical	Standard deviation	2 (7.7)
M_{11}	Mathematical	Frequency distribution	13 (50)

Note. * Students who drew boxplots calculated the three quartiles of the distribution of scores. Quartiles are also mathematical models.

Table 2: Summary of the different types of models developed by the participants in this study

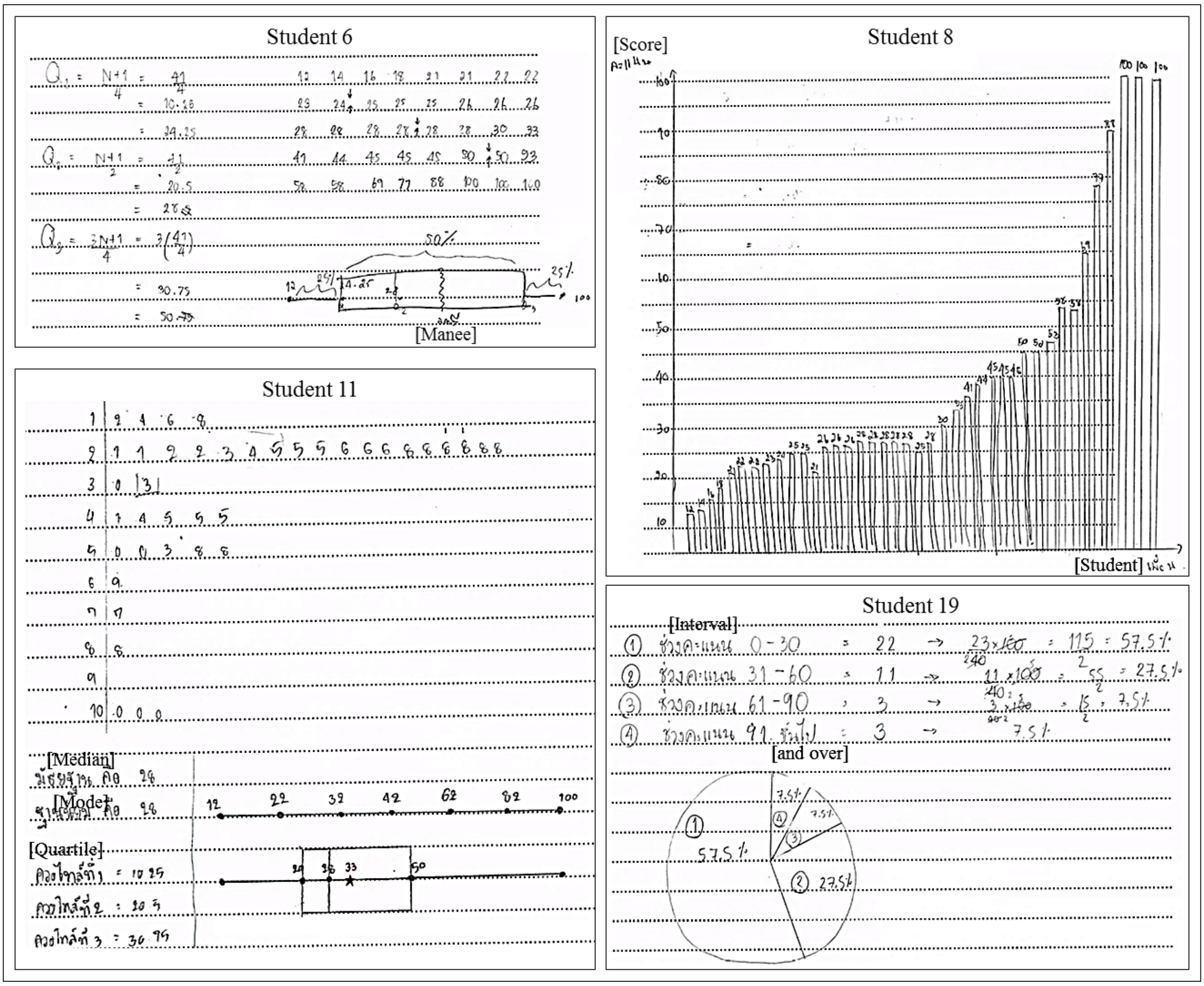

Figure 2: Examples of different types of models developed by the participants in this study

On average, two to three models were provided per student, being the mode 2 models (9 students, 34.6%), and the number of models developed by participants ranged from 1 to 5. As can be seen in Table 2, sorted data arrays (13 students, 50%) and frequency distributions (13 students, 50%) were the models most used by students when answering the given task. In relation to students' reasoning about variability in this task, it is worth noting that only 1 student out of the 14 who provided evidence of either value-relation reasoning or magnitude-comparative reasoning drew and used a boxplot to support her argument. This comes as no surprise, given that value-relation reasoning and magnitude-comparative reasoning focus on attending to individual or groups of cases, rather than aggregate features of data sets as a whole (Watson, 2018). On the contrary, 6 out of 7 of the students who drew a boxplot in this task gave reasons explicitly using proportions (i.e., explicit proportional or distributional reasons). Moreover, all the participants engaged in some total ordering of all sample elements, either via a sorted data array, a frequency distribution, or a stem-and-leaf plot.

Some of the models developed by participants in this study are shown in Figure 2. Some misunderstandings in the construction and use of certain models were identified during the data analysis (e.g., the bar graph in Figure 2). This issue will be addressed in a future publication.

Conclusions

It can be concluded that participants relied more on value-relation or magnitude-comparative reasons in their arguments, in comparison to distributional ones. However, all participants used, implicitly or explicitly, proportional reasoning, which was evident from the use of proportions or relative frequencies in all their responses. This fact suggests that participants are somehow good proportional reasoners in general, but still weak distributional ones. The source of this difficulty could be rooted on the fact that, in Thailand, students learn about stem-and-leaf plots, boxplots and quartiles in Grade 12 (participants' current grade), and they may need more time and/or experience to master those ideas.

It can be also concluded that asking students to provide persuasive arguments based on data, in the context of a socially open-ended problem, will give them opportunities to, among other things, (1) make meaningful use of distribution features such as center, proportions and spread; (2) structure variation and coordinate variability and chance by engaging actively in modeling challenges; (3) develop an aggregate view of data; (4) engage in data-driven argumentation by enacting the practice of argumentation as advocacy; (5) actively engage in the decision-making process, in order to choose and generate models able to concisely describe a body of data and capture the essential characteristics of a real-life phenomenon; and (6) develop students' awareness of variability.

Finally, we can conclude that socially open-ended problems with the characteristics described here seem to be a way to help teachers achieve the aims of the mathematics curriculum about statistics education. Moreover, the results reported here provide teachers and researchers with insight into students' ways of reasoning about and modeling variability, which can be used to improve teaching practice, design better curriculum materials, and uncover misconceptions about statistical ideas.

References

Becker, J., & Shimada, S. (1997). *The open-ended approach: A new proposal for teaching mathematics*. Reston, VA: NCTM.

Blair, J. A. (2003). Relationships among logic, dialectic and rhetoric. In F. H. van Eemeren, J. A. Blair, C. A. Willard & A. F. Snoeck Henkemans (Eds.), *Anyone has a view: Studies in argumentation* (pp. 91–107). Dordrecht: Kluwer Academic Publishers.

Gal, I. (2004). Statistical literacy. Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield, *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht: Kluwer Academic Publishers.

Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning*. New York: Springer.

González, O., & Chitmun, S. (2015). Secondary teachers' pedagogical content knowledge to promote students' data-based decision-making ability: The cases of Japan and Thailand. In C. Vistro-Yu (Ed.), *Proceedings of the 7th EARCOME* (pp. 475–484). Quezon City: MATHTED.

González, O., & Chitmun, S. (2017). Using sport-related socially open-ended problems to coordinate decision-making, variability modeling and value awareness. In T. Dooley & G. Gueudet (Eds.), *Proceedings of the 10th CERME* (pp. 780–787). Dublin: Dublin City University and ERME.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, *4*, 635–679.

Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching statistics in school mathematics* (pp. 347–358). New York: Springer.

Mulligan, J., & English, L. (2014). Developing young students' meta-representational competence through integrated mathematics and science investigations. In J. Anderson, M. Cavanagh, & A. Prescott (Eds.), *Proceedings of the 37th Conference of MERGA* (pp. 493–500). Sydney: MERGA.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning*, *5*(2&3), 131–156.

Royal Thai Ministry of Education (MOE) (2008). *The basic education core curriculum*. Bangkok: Bureau of Academic Affairs and Educational Standards.

Shimada, I. (2015). *Nurturing the competence of being engaged with diverse values in mathematics education: An effort through solving socially open-ended problems* (in Japanese) (Unpublished doctoral dissertation). Hiroshima University, Hiroshima, Japan.

Shimada, I., & Baba, T. (2015). Transformation of students' values in the process of solving socially open-ended problems. In K. Beswick, T. Muir, & J. Wells (Eds.), *Proceedings of the 39th Conference of the IGPME* (Vol. 4, pp.161–168). Hobart, Australia: PME.

Shaughnessy, J. M., Ciancetta, M., & Canada, D. (2004). Types of student reasoning on sampling tasks. In M. Høines & A. Fuglestad (Eds.), *Proceedings of the 28th Conference of the IGPME* (Vol. 4, pp. 177–184). Bergen, Norway: Bergen University College.

Watson, J. M., Fitzallen, N., Fielding-Wells, J., & Madden, S. (2018). The practice of statistics. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics* (pp. 105–138). Cham, Switzerland: Springer.