Introducing Statistical Variation in Year 3 in a STEM Context: Manufacturing Licorice

Jane Watson[a], Noleine Fitzallen[a], Lyn English[b], Suzie Wright[a]

[a]College of Arts, Law, and Education, University of Tasmania, Hobart, Australia;

[b]Faculty of Education, Queensland University of Technology, Brisbane, Australia;

As STEM education becomes more prominent at the school level, many possibilities could be suggested for beginning experiences to create a foundation for young students as they begin their journeys with respect to STEM learning. The research reported here is based on the premise that building statistical understanding will enhance young students' learning as they engage with STEM learning experiences. This paper hence reports on an activity created to introduce students in Year 3 to the concept of statistical variation with data in a STEM-related context where variation occurs in an easily measured and realistic fashion. Students' capabilities to appreciate the fundamental nature of statistical variation and use it for comparison in a STEM context were assessed through responses in student workbooks, to questions on an end-of-year survey, and in individual interviews. The results illustrate the beginning stages of student thinking about variation and the use of data when learning is embedded within a STEM context. They show that students can take on the idea of variation and use it in explaining their experiences with the hands-on activity, which involved comparisons of a hand-made and machine-made product.

Keywords: variation, data modelling, STEM, manufacturing, Year 3

1

The research reported in this paper is the first in a longitudinal study of primary students' experiences of modelling with data in STEM contexts. Beginning with the suggestion that it is possible to link statistics and STEM education, the fundamental concept of variation is introduced as a foundation for statistical understanding, and hence its contribution to STEM investigations. The usefulness of data modelling for answering questions in the STEM disciplines at the school level is also discussed. The STEM context for the Year 3 students' investigation was manufacturing a product using a simple machine and by hand. With the goal of producing the same product in two ways, students explored the variation within and between the two methods. The general research question hence became: What capabilities do Year 3 students demonstrate with respect to developing statistical understanding in relation to variation, through data modelling experiences set within a STEM context?

STEM education and statistics

The growing focus on STEM in countries around the world to advance sustainable economic development [e.g. 1–3], has created the need for corresponding interest in "STEM education" [e.g. 4–5]. The phrase STEM education immediately raises debate within the community of educators and researchers interested in STEM education about the degree of integration of all four disciplines required for any given activity planned for students [e.g. 5–8]. Nonetheless, it has become accepted that the integration of STEM can manifest in different combinations of connections made among two or more of the four disciplines [9–10], and a number of different models have been proposed [e.g. 8,11–12]. Common to the varying perspectives and models is the utility of STEM learning experiences to support students to make connections to real-world contexts and apply problem-solving strategies. However, more attention to the content knowledge that supports STEM disciplines within the relevant contexts is advocated [7,13].

A perspective useful for the framing of the research reported in this article is that of Shaughnessy [14,p.2] as President of the National Council of Teachers of Mathematics in the United States who claimed,

STEM is an amalgamation of very complex and intertwined scientific disciplines. STEM is not a content area in and of itself, and we should not talk about it as if it were. Those who are implementing any STEM program should be able to identify the specific mathematical foci of the program.

Fitzallen [7,p.242] added support to this view when she stated "mathematics should be given more standing and be considered an enabler or imperative for the advancement of understanding of concepts in other disciplines." Taking Shaughnessy's advice, the stance of this research is to introduce the fundamental concept of variation as the "specific mathematical focus" of data modelling opportunities that arise naturally from a meaningful STEM context.

In relation to choosing variation as the mathematical focus of this project, the question might be posed as to why statistics has the potential to provide the links among the STEM disciplines. Certainly, Watson [15] illustrated the potential connections across the disciplines when she explored the links among statistics, and mathematics and science curricula from three countries: Australia, New Zealand, and the United States of America. In addition, the *New Generation Science Standards*, which incorporate big ideas from science, technology and engineering and link them to mathematics and literacy [16], provide instances where data are instrumental for science and engineering practices at every level of the curriculum. Indeed, there is a synergy between the STEM disciplines and the application of statistics. The STEM contexts provide meaningful situations where statistical understanding can be developed. After all, Rao [17, p.152] asserts:

Statistics ceases to have a meaning if it is not related to any practical problem. There is nothing like a purely statistical problem which statistics purports to solve. The subject in which a decision is made is not statistics. It is botany or ecology or geology and so on.

This background provides the motivation for initiating young children's experiences with both STEM and data modelling in a straightforward familiar context with a "natural" hands-

on activity and the requirement for representation and communication. The project not only contributes to the burgeoning field of research in STEM education in relation to the early years of primary education but also addresses calls for STEM education to "feature discussion, visualization, and other forms of representation (e.g., drawing, writing, graphing) to promote learning that leads to generalization of important concepts and practices" [18,p.2].

Variation

The most fundamental idea underpinning statistical investigations is that of variation. Indeed, without variation, there would be no need for Statistics. Cobb and Moore [19,p.801] made the often-quoted statement that "[Statistics] exists not for itself but rather to offer to other fields of study a coherent set of ideas and tools for dealing with data. The need for such a discipline arises from the *omnipresence of variability.*" Shaughnessy [20] suggested that although statisticians realized this, early statistics curricula were procedure-based and focused first on centres (expectation), measured for example by the mean, rather than variation, measured by the standard deviation.

Only since the content of curriculum documents has begun to suggest other ways of considering variation in data, for example by observing the range or other visual attributes of graphical representations, have researchers at the school level begun to carry out research on students' developing understanding of this fundamental concept. Many science contexts have been used and age groups studied over the intervening years; for example, Lehrer and Schauble [21] considered variation in the context of plant growth in Year 5, Shaughnessy and Pfannkuch [22] reported on the reactions of high school students studying the wait time for eruptions of Old Faithful, and Watson and Kelly [23] explored variation in the weather with students from Year 3 to Year 9. As well, research began to consider the relationship of the development of variation and expectation [24–25]. Based on this research, Watson [26]

4

suggested that appreciation of variation arises for children before the appreciation of expectation.

In relation to students in Year 3, Watson and Kelly [27] reported on the development of student understanding of variation after experiencing 10 lessons based on hands-on activities to build and reinforce understanding of basic chance and data concepts, particularly the ideas associated with variation. The activities were based on contents of small Smarties^{TM} boxes, size of families, chance devices, sampling objects from opaque bags, standing on one foot with eyes closed, and blowing pencils across the floor. Using pre- and post-tests with items from Watson and Callingham [28], students showed significant gains overall, and on items related to variation in context.

More recent research has moved directly into the classroom to document children's capabilities to take on statistical understanding as lessons occur and students interact with data through statistical investigations [e.g. 29]. As part of developing informal inference [30], English and Watson [31] introduced a measurement activity in Year 4 classes, where all students in the class measured the arm span of one student and created a representation of the class data by hand. All students then each had their own arm span measured once. All data were placed in *TinkerPlots* [32] files and students were asked to consider the variation within each data set and also to compare the variation between their two representations. The reasons for the differences involved understanding the contexts for the two data collections, applying understanding from Science, and using Technology for their analyses. The research with these students was extended in Year 6 to consider both variation and expectation explicitly in the context using *TinkerPlots* technology to toss a virtual die [33].

Although there has now been considerable research on variation in many different contexts for various year groups, the research reported here extends previous research in three ways. First, in the spirit of building understanding of STEM concepts, the context

chosen was manufacturing, where variation can be easily observed and described by Year 3 children in relation to a beginning appreciation of engineering and technology, and what engineers do. Second, it created the need for data and data modelling to understand how a STEM-created product was different from a hand-made product. Third, it provided the opportunity at this level to explore the early stages of development that involve comprehending, describing, representing, predicting, and retaining ideas about the fundamental notion of variation. As well, the study reinforced the principles for young learners of the Early Childhood STEM Working Group [18] related to developing "natural" STEM inclinations and focussing on representation and communication in doing so.

Data modelling

An early description of Lehrer and Romberg [34] relates the use and construction of data to data modelling and to mathematical modelling. They suggest, "The very idea of data entails a separation between the world and a representation of that world" [34,p.70]. As with mathematical models, data models provide symbolic structures that can be manipulated. "For example, by manipulating data, new questions can be posed about relations among elements of the data structure" [34,p.70]. New instances of the model can be generated and further inferences made that may be applied to the real world. Growing from Lehrer and Romberg's [34] attention to data modelling are many descriptions of the relationship of models and modelling for solving problems involving data, for example, recycling and the environment [35–36], traffic congestion [37–38], plant growth [21], and designing rockets [39]. Crites and St. Laurent [35] go further to add that an essential outcome of modelling is the understanding that statistical models are useful for describing data and making decisions from data. Lehrer and English [40] consolidate recent research specifically on children's modelling of variability.

At the school level, data modelling is often identified with "statistical problem-solving", as proposed by the American Statistical Association's *Guidelines for Assessment and Instruction in Statistics Education (GAISE) Report* [41]. That document outlines a model for statistical problem solving across three developmental levels of the years of schooling. In outlining the statistical problem-solving process *GAISE* proposes four components. These are presented in Figure 1. The elaboration of each of the components focusses on variability: anticipating, acknowledging, accounting of, and allowing for it. Accepting this framework, STEM contexts should provide excellent environments within which to model statistical problem-solving. For students who are being introduced to statistical problem solving for the first time, it would not be reasonable to expect them to implement the entire *GAISE* model by themselves. In this study half of the steps were introduced by the teacher in discussion with the class: clarify problem, formulate question, design data collection, and select appropriate methods. The students then completed the other half: collect the data, analyze the data, interpret the analysis, and relate to the original question.
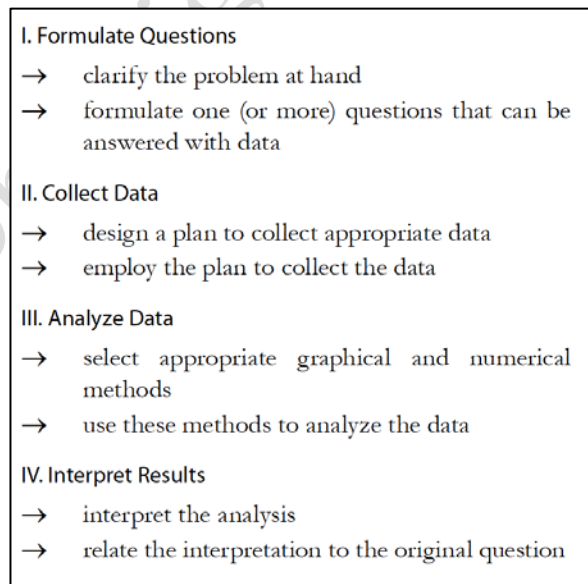

I. Formulate Questions
→ clarify the problem at hand
→ formulate one (or more) questions that can be answered with data

II. Collect Data
→ design a plan to collect appropriate data
→ employ the plan to collect the data

III. Analyze Data
→ select appropriate graphical and numerical methods
→ use these methods to analyze the data

IV. Interpret Results
→ interpret the analysis
→ relate the interpretation to the original question

Figure 1. GAISE model for statistical problem-solving [41,p.11].

Significant in the decision-making process based on data modelling is the acknowledgement of uncertainty. Although professional statisticians acknowledge

uncertainty on theoretical grounds with formal inferences based on p-values and confidence intervals, Makar and Rubin [30] suggest an informal approach to decision-making at the school level. At this level a statistical question is answered for a population based on evidence from a sample, acknowledging a degree of uncertainty. The degree of uncertainty depends on the quality of the evidence, which is related to the usefulness of the model created for the context from the data and the data's variation [35].

Research approach

The activity reported in this article was the first, as part of a longitudinal project following the students from mid-way through Year 3 to the end of Year 6. With the aim of designing and implementing classroom activities using data modelling experiences to facilitate STEM learning, the first activity sought to introduce the fundamental concept of variation in a hands-on context that could be linked to an authentic STEM-related context outside of the classroom. The research was set within an interpretive paradigm and employed predominantly qualitative research strategies [42] to establish the breadth of the students' capabilities to describe and apply notions of variation and data modelling.

Participants

Seventy Year 3 students in three classes in two schools in different Australian cities completed the activity that is the focus of this study. Their mean age was 9.0 years (range 7 years 8 months to 9 years 7 months). The schools were parochial schools, one a girls' school ($n = 22$), and one of mixed gender ($n_{male} = 29$, $n_{female} = 19$). Overall the gender balance across the schools was 41% male and 59% female. Permission was given by parents for the students' data to be included in the study.

Procedure: Manufacturing Licorice activity

8

Given the experiences of Watson, Skalicky, Fitzallen, and Wright [43] with activities based on using the Play-Doh[TM] Fun Factory (see Figure 2), an extruder toy, in professional learning sessions with teachers and with children in Years 1 and 3, it was decided that this context would be meaningful to Year 3 students in terms of looking at the variation resulting from creating a product by hand or by machine. The manufacturing context provided a link to Engineering and Technology, whereas measuring attributes of the product linked to the measurement area of Mathematics and to the need for fair tests in Science inquiry skills [44]. The product chosen for the manufacturing context was licorice because there was a product on the market with licorice sticks that were 1 cm in diameter (see Figure 2), the same as the diameter produced by the Play-Doh[TM] extruder. The length of the commercial sticks was approximately 8 cm and hence students could be asked to create their own facsimile licorice by hand and then using the extruder. Earlier work with middle school students creating 5 cm sausages with the same device [45–46] was the motivation for the activity but the context was changed to licorice for ease of stimulating interest in the activity by showing an accessible commercial product of the type to be produced by the students.

Figure 2. Play-Doh[TM] Fun Factory extruder and the commercial licorice.

Implementation

The regular classroom teachers implemented the activity and facilitated the classroom discussions. In the first lesson, the students were introduced to the activity by the teacher

asking questions about products made by factories that can also be made by hand, leading to the ways they might be different. Many examples were suggested by the students, for example, chocolate, furniture, bread, clothing, toys, lollies, woollen jumpers, and wooden spoons. A discussion followed about machines being designed by engineers, and technology ensuring consistency in the products. A packet of licorice was shown to ensure that all students were familiar with the product. This was followed by showing two videos on the production of licorice, one by the manufacturer of the product on display (www.facebook.com/riccilicorice) and another with more details on the manufacturing process (www.youtube.com/watch?v=AVAnYMVRmdI). Again the consistency of products produced on the videos was discussed after each video was shown.

Students were then asked to think about how consistent they could be in making licorice by hand and were challenged to be as "exact" as possible using Play-Doh^{TM}. Students in groups of three were given Play-Doh^{TM} of three different colours, rulers, scales, pencils, and workbooks to make their licorice sticks (see Figure 3). There were reminders on reading the rulers and scales, on the specifications of 1 cm diameter and 8 cm length for the sticks, and on carefully recording the mass (g) on a sticky note for each licorice stick produced. Each student made three licorice sticks, weighing and recording the measurements of mass in their workbooks. When finished making the licorice sticks, students recorded in their workbooks the mass of the lightest piece, the heaviest piece, and the difference between the two, followed by giving three reasons why the masses may not have been the same. The next task was to draw a representation of their group's data in their workbooks. No constraints were placed on this representation, with students encouraged to tell the story of their group's data in any way they liked. Groups were then asked to share their representations with the rest of the class and explain what was shown about the data collected.
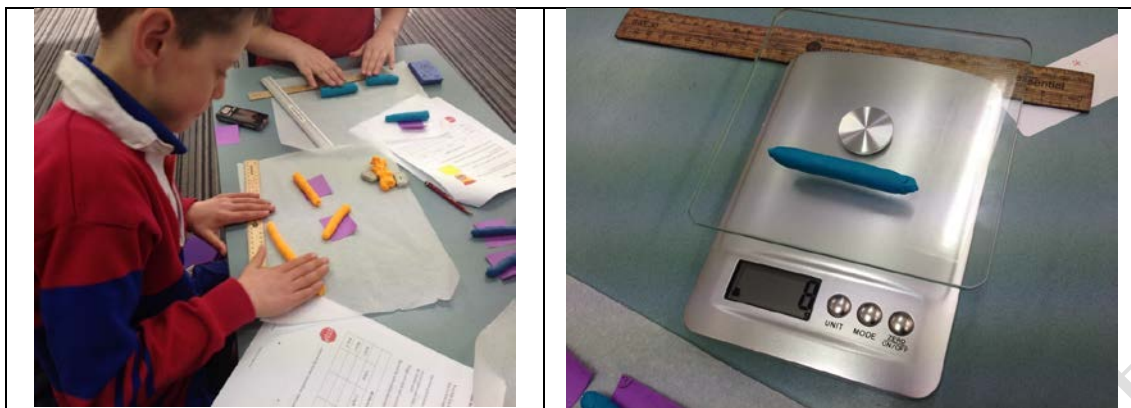
Figure 3. Handmade sticks and scales.

Focusing on the differences across the groups, the teacher moved to combining the groups' results for the entire class to consider the overall consistency. The teacher drew a horizontal axis on a large sheet of paper attached to the wall and suggested students record their data using their sticky notes. There followed discussion about the "ends" of the axis line and checks were made around the class to find the lightest mass and the heaviest mass, which were used to determine the range of the data. The teacher then labelled the axis and students each chose one of their sticky notes and placed them on the axis according to the mass recorded. The plot created by one class is shown in Figure 4. The teachers reported that the students had not been exposed to frequency plots like this before creating the plot.
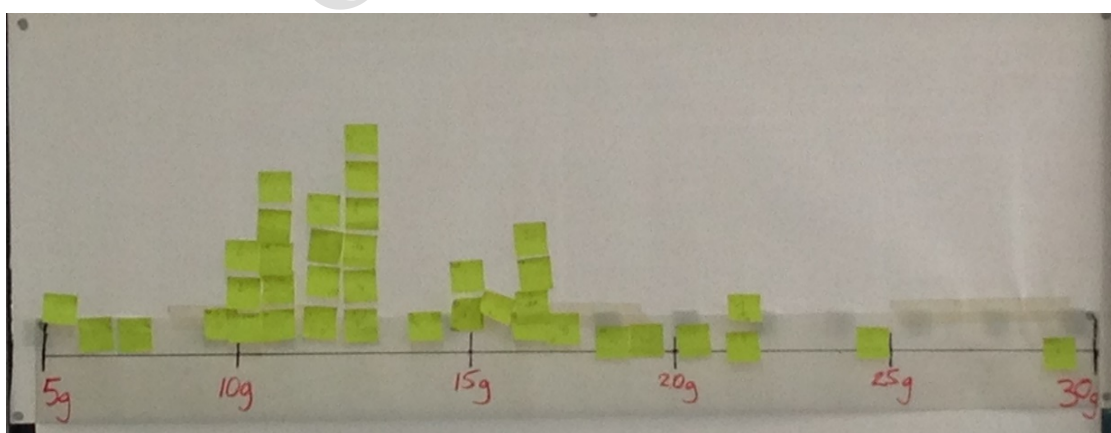

Figure 4. Class plot of masses of hand-made licorice sticks.

A series of questions followed to focus on the variation in the data, although the word was not introduced at that time. Students described shapes they could see in the class data, for

example gaps, and identified clusters. Students were then asked to suggest typical values and to make predictions in their workbooks about the mass of another licorice stick if they made it themselves or if another student came into class and made one. A few students were asked to fill in sticky notes with their predictions and place them on the plot on the wall, explaining the reason for the value selected. The first lesson closed with an introduction to the concept of variation, including the word, with a series of questions about the differences in the students' data, why it occurred, and how it might be different if the students made their licorice sticks with a machine rather than by hand.

In the second lesson, students were shown the Play-Doh^{TM} extruder (Figure 2), with a demonstration of how to use it and a reminder of the specifications for the licorice sticks of 1 cm diameter and 8 cm length. Each group of three students had an extruder to use. Other instructions were the same as previously with each student making three licorice sticks and recording the masses on sticky notes (see Figure 5). Again students were asked to create a representation for their group's data and choose one (or more) of their results to place on another horizontal axis where the teacher had labelled the endpoints of the axis with the values of the endpoints from the hand-made licorice sticks.

Figure 5. Manufacturing licorice sticks with the Play-Doh^{TM} Fun Factory extruder.

The machine-made plot created by the same class as in Figure 4 is shown in Figure 6. The two plots were viewed side-by-side on walls in the classroom. Students then answered the same questions as before in their workbooks about the class plot for the machine -made licorice sticks and then about comparing the two plots. They were also asked which kind of licorice sticks they would prefer if the product were real. Finally, they were asked to draw a picture and write a story about the investigation undertaken.
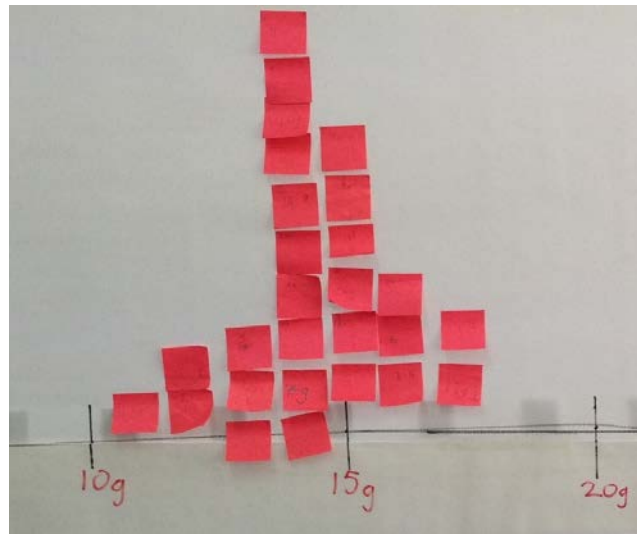

Figure 6. Class plot of masses of machine-made licorice sticks.

At the end of the school year, four months after the activity, 67 students answered four questions about the activity on the survey administered as part of the longitudinal evaluation of the project. The survey included 18 other questions related to statistical literacy, used in previous research [e.g. 28]. Also in the girls' school, all 22 students who completed the activity were interviewed individually on the activities during the year, including this one. The interview was structured around the plots of the hand-made and machine-made sticks shown in Figures 4 and 6 from the other site. Students were taken through a series of questions reflecting the progress of the activity. They were asked to describe the shape of the plot in Figure 4 and what they could tell about the licorice sticks made by the other class. Prompts were given for an explanation of "variation". They were then shown Figure 6, with

the same questions, followed by questions on the difference between the two types of licorice sticks. Finally they were asked what they had learned from the activity.

Data collection instruments

A variety of data collection instruments were utilised to capture the students' thinking and reasoning as they worked though the Manufacturing Licorice activity, as described in the previous section. Video was used to record the classroom discussions and those of six groups of three students while making the licorice sticks.

To maintain the students' anonymity, students' names used in the interviews and written on the workbooks were replaced with unique student codes, for example, ID465. The student codes are used to identify the student work presented in this article. It was not always possible to identify the students speaking during the classroom discussions. Therefore, excerpts of transcripts of the classroom discussions use student numbers (e.g., S8) to identify the different students in each of the excerpts.

Specific objectives of the data analysis

In relation to the general research question on the capabilities of Year 3 in developing statistical understanding, particularly in relation to the fundamental concept of variation, using a STEM context, the following specific objectives were the focus of the evidence collected in the Manufacturing Licorice activity. They point to the expected starting points for young students.

Objective 1

Report on Year 3 students' capabilities to comprehend, describe, and represent the variation experienced in making the licorice sticks by hand and by machine.

Objective 2

Report on the ways Year 3 students use their notions of variation to suggest typical values in a data set and to make predictions.

Objective 3

Report on Year 3 students' capabilities to describe the variation within the two

class plots and the reasons for the differences in production by the two methods,

and to justify a preference for hand-made or machine-made licorice sticks if they

were real.

Objective 4

Report on Year 3 students' retention of the concepts introduced in the

Manufacturing Licorice activity four months later.

Data analysis

Following a preliminary analysis of the data from the girls' school [47], the data used in the

final analysis across sites are the responses written in the students' workbooks, transcripts of

all classroom sessions, the responses in the end of year survey questions, and transcripts of

individual interviews of one class. The questions and rubrics for the workbooks are presented

in Table 1 and for the survey questions in Table 2. Because the students were being

introduced to a new concept, variation, through examples of its occurrence, it was of interest

to classify their capabilities in describing their new understanding in the STEM context. The

coding of responses was hierarchical in terms of the recognition of the requirements of the

particular question. At times responses reflected single or multiple observations of the

context of the activity. These responses could be assessed with the Structure of Observed

Learning Outcomes (SOLO) model [48] as children transition into the concrete symbolic

mode, because it classifies responses by how many of the elements required for the task are

combined and how the process is described. For Iconic responses, no relevant elements are

employed and the response may be idiosyncratic or out of context (Code 0). For Unistructural

responses, a single element is used and contradictions in claims made may not be recognized

(Code 1). For Multistructural responses, several elements may be employed, usually in a

sequence without closure, depending on the question (Code 2). For Relational responses, several elements are related together to achieve closure for the question (Code 3). The elements that were available for use in the Manufacturing Licorice activity were the raw data collected for the masses of the licorice sticks, the mechanism/procedure for creating the licorice sticks, the characteristics of the representations created by the students, and the characteristics/descriptors of the class plots representing the two sets of data. For various questions in the workbook, it is how these elements were combined to address the focus of the particular objective, for example, variation, prediction, and differences between the two types of licorice sticks, which determined the level of response. Because this activity was the students' first formal encounter (as confirmed by the teachers) with the concept of variation and the necessity to explain the phenomenon in words, many of the questions in the workbooks did not require Relational responses. In one case, however, some responses at the Relational level went further to incorporate the precise representation or language of variation; these responses were coded 4. In the case of the representations created for the group's licorice sticks and in telling the story of the activity, types of representations were of particular interest because the students had had different previous experiences with graphing; the girls' school students had been taught specifically how to create bar (value) plots, whereas the other students had not. The classification of *types* of graphs, hence, was not hierarchical and included Scaled Graph, Table or Tallies, Pictograph, or a drawing that could not be categorized. The representations were also analysed according to the SOLO model, in terms of the representation of the amount of raw data displayed (elements). The codes in Tables 1 and 2 greater than 1 reflect the SOLO levels as described in this paragraph. Examples of the responses are provided in the Results.

All written responses of students were entered into spreadsheets with drawings scanned and filed by student ID. The fourth author and an experienced researcher coded all

responses independently with an agreement rate of 94%. All discrepancies were resolved by discussion between the two.

The first three questions in the workbook (Q1 to Q3) asked students to record data from weighing their hand-made licorice sticks, the lightest and heaviest sticks, and the differences between the lightest and heaviest. The same questions were asked for the machine-made sticks. These were not coded but the accuracy is reported in the Results.

Table 1. *Questions and rubrics for the workbook questions about the two ways of making the licorice stick (Q4 to Q9 asked for both hand-made and machine-made).*

Question	Level	Level descriptor
Objective 1: Comprehending, describing, and representing variation		
Q4: Give 3 reasons why all the pieces may not have been the same. (Reasons must be different)	2	A reason that implies comparison with words related to the creation of the pieces other than heavy/light/weight.
	1	A description of difference without a reason; mentions other characteristics not associated with weight (e.g., twistiness, colour, cracks).
	0	No Response [NR]; idiosyncratic.
Q5: Use the space below to represent your group's data. You may do this any way you like.[1]	2	Representation shows or records all the data in a fashion that can be related to the activity.
	1	Incomplete (missing data)/unclear representation.
	0	Unintelligible representation
Q6: Describe the shape of the class plot.	3	Description of multiple characteristics that include appreciation of variation.
	2	Description of a single characteristic of data in graph that displays appreciation of variation.
	1	Description of a characteristic of plot but not acknowledging variation.
	0	NR; refers to "picture" not data.
Q7: What does the shape of the plot tell you about the variation in the licorice sticks made by the class?	2	Refers to specific change or comparison of difference.
	1	Refers to a single source of difference without description of what it is.
	0	NR; idiosyncratic; out of context.
Objective 2: Identifying typical values to make predictions		
Q8: How would you describe the typical mass for the sticks made by the class?	1	The midpoint or a range including the midpoint.
	0	Very large value; vague description.
Q8 Why: Why do you think this might be the case?	1	Reason associated with the closeness of the values on the class plot.
	0	Other reason.
Q9: If you made one more piece of licorice, what do you think (predict) its mass might be?	1	Any values within the range of masses for the class.
	0	Other value, e.g., 100g.
Q9 How: How did you decide?	3	Reason based on middle/mode of class data summarising the data in the class plot.
	2	Reason based on students' personal data.
	1	Non-statistical reason related to the data.
	0	NR; undecipherable.

[1] See also classification of the Type of representation in the Results.

Objective 3: Comparing plots to make justifications		
Q11: List the differences between the two plots (hand-made and machine-made).	3	More than one difference including specific comparison.
	2	One difference, with a specific comparison.
	1	Implied comparison without mention of the two data sets.
	0	Variation not present.
Q12: What does this tell you about the two ways of making licorice sticks?	2	Notes consistency: Machine-made more "the same" than hand-made.
	1	Difference but not related to the making; relates to making licorice but no mention of difference.
	0	NR; response not related to ways of making.
Q13: Which type of licorice would you rather buy if they were real? Why?	2	Statistical response: Machine-made, based on consistency.
	1	Non-statistical response; personal preference.
	0	NR; choice with no reason; idiosyncratic.
Q14 Picture: Draw a picture that shows what you found out in the investigation.	4	Data representing two ways of making licorice sticks shown: Hand-made data displays *more variation in mass* than machine-made.
	3	Data representing two ways of making licorice sticks shown; difference is implied but not clearly visible, or not related to the actual data collected.
	2	Two ways of making licorice sticks shown pictorially; no further representation discernible in relation to data or variation.
	1	One way of making licorice shown, with some type of representation of data; variation may be internal to that way of making.
	0	NR; one way of making licorice sticks with no realistic representation of the output/data.
Q14 Story: Write a story to match your picture.	4	Two ways of making licorice sticks mentioned/described, with specific comment that hand-made had *more variation in mass* than machine-made.
	3	Two ways of making licorice sticks mentioned/described; difference is implied but not clearly described, or not related to the actual data collected, as in *easier*.
	2	Two ways of making licorice sticks mentioned only; no further discussion discernible in relation to appropriate data or variation.
	1	One way of making licorice described, with some type of mention of data; variation may be internal to that way of making.
	0	NR; one way of making licorice sticks with no description of the output/data.

Table 2. *Questions and rubrics for the end-of-year survey about the two ways of making the licorice stick.*

Objective 4: Retaining concepts introduced in the licorice activity (Surveys)		
Question	Level	Level descriptor
Explain how we collected the data for the activity.	3	Description includes making the sticks in two ways – by hand and by machine – and using the scales to weigh them.
	2	Description two aspects of the activity, e.g., making (perhaps implicit) and weighing the sticks; or both ways of making the sticks but no mention of weighing them.
	1	Description of a single part of the activity in general terms, e.g., making licorice sticks, measuring the length or other characteristics.
	0	Description not related to data collection; NR.
What data did we collect/and record?	2	Specific reference to mass of sticks from hand-made and machine-made licorice sticks.
	1	Reference to mass of sticks as data in general/vague terms or to two methods (only).
	0	Reference to sticks and scales but not data; idiosyncratic response; NR.
Sketch two graphs to show the data we collected.	3	Two graphs showing the difference in shape (variation) and labelled appropriately.
	2	Two graphs showing difference in shape (variation) without measurement scale or without complete data.
	1	Only one graph, perhaps with inappropriate scale.
	0	Idiosyncratic graph not related to data; NR.
What did we conclude from our investigation?	3	Describes difference using statistical language, such as spread, range, clustered/clumped, gaps, consistency, variation.
	2	Describes difference in colloquial terms; sense of variation with unsubstantiated conclusion.
	1	No reference to difference in methods; personal opinion.
	0	Idiosyncratic response; NR.

Extracts from the transcripts of the classroom discussions are presented in relation to Objectives 1, 2, and 3, to reinforce aspects of the responses recorded in student workbooks. As a measure of the degree of recall from the activity to the end-of-year survey, the Pearson product-moment correlation coefficient was calculated between the sum of scores for Objectives 1–3, associated with the in-class experiences, and Objective 4, related to the end-of-year survey.

For the end-of-year interviews in the girls' school, the familiarity from the classroom experience of the interviewer and the individual students meant some girls naturally volunteered more information and hence it was not possible to code responses hierarchically. The interview transcripts were read by two authors with responses to the first question, "Describe the shape of the plot (in Figure 4)" and the last question, "What did you learn from this activity?", categorised based on the rubric in Table 3. As well, if not mentioned by the students, the interviewer specifically asked about the variation in the licorice sticks. Questions about the shape of the second plot (Figure 6) and comparison with the first were not coded because often teaching took place between the specific questions on the shape of the plots. The overall agreement between the two authors on coding was 86% for the categories selected with discrepancies resolved after discussion.

Table 3. *Rubrics for interviews.*

Objective 4: Retaining concepts introduced in the licorice activity (Interviews)		
Question	Code[1]	Code descriptor
Based on plot for hand-made sticks (Figure 4): Describe the shape of the plot.	A	Student used language about the appearance of the heights of the sticky notes in the plot.
	B	Student discussed the spread of the data.
	C	Student discussed the typical values or most being in the middle.
Explicit understanding of Variation	Yes	
	No	
What did you learn?	D	Student did not know or just "had fun".
	E	Student suggested learning "new words" without further description.
	F	Student noted learning a technique (e.g., weighing or measuring).
	G	Student mentioned learning about variation or described the concept.

[1]Codes are for non-hierarchical categories of response.

Results

Objective 1: Comprehending, describing, and representing variation

Because the questions in the students' workbooks were identical for the two data collections in the activity, the results for Objective 1 are reported together. Students recorded the masses of their three licorice sticks each time (Q1) and to reinforce their basic arithmetic skills they were asked to record the lightest and heaviest pieces for their group (9 values) (Q2), and then to find the difference (or variation) between the lightest and heaviest (Q3). Masses recorded between 5g and 30g were considered reasonable for the activity. As seen in Table 4, most students had no difficulty distinguishing the lightest and heaviest masses for their group and the difference for the range. A few gave a physical description of the difference such as "the lightest one was thin[n]er and twisted" (ID121) or "diff[e]rent colour and diff[e]rent mass" (ID125).

Table 4. *Reporting on data collection.*

	Hand-made	Machine-made
Reasonable values (Q1)	90%	96%
Lightest (Q2)	99%	97%
Heaviest (Q2)	99%	97%
Difference (Q3)		
Numerical	96%	91%
Qualitative	3%	6%
Missing	1%	3%

When asked to give three reasons why the licorice sticks were not the same (Q4), examples of the reasons are given in Table 5. The Code 1 responses focused on the mass again or a description of another feature of the sticks that was not related to mass, such as colour. The Code 2 responses provided a second attribute that could contribute to the difference in masses of the licorice ticks. Examples are in the table. Across the three reasons for the hand-made sticks, one student did not record a response, and three gave only

idiosyncratic responses (Code 0) (6%). Of the 66 who gave at least one non-idiosyncratic

response (Code 1 or 2), 94% were coded 2 (Multistructural) for at least one of the responses,

and 30% provided all three Code 2 responses. Across the three response for machine-made

sticks, 3 students did not record a response or gave only idiosyncratic responses (4%). Of the

67 who gave at least one non-idiosyncratic reason, 91% were coded 2 for at least one of the

responses, and 31% provided all three Code 2 responses. Over all of the 417 responses given

as the reasons for variation within the hand-made and machine-made sticks, 60% were Code

2, 22% were Code 1 and 18% were either idiosyncratic or a missing response where others

were at a higher level (Code 0).

Table 5. *Responses for reasons for differences in the licorice sticks, Q4, (percentage of all responses given).*

Level	% Hand-made	Examples Hand-made[1]	% Machine-made	Examples Machine-made
2	61%	Some are thin and some are fat. (ID125) difrent [different] peple [people] mack [make] them (ID011)	59%	Maybe the pieces were not measured cor[r]ectly. (ID122) They might operate the machine differently (ID124)
1	21%	the wight [weight] is differnt [different] (ID008) Some are twistier (ID102)	23%	Different mass. (ID134) Some mite [might] have had cracks. (ID119)
0	18%	They might of all not been exzaclly [exactly] the same (ID013) We had difrent idea's [different ideas] (ID132)	18%	Very hard to tell. (ID103) It is just a play-doh michiane [machine] not a licorice michiane [machine]. (ID129)

[1] Spelling corrections have been put in square brackets and student codes in parentheses.

The representations drawn for the hand-made licorice sticks (Q5) reflected very much

the previous experience of the classes with graphing. More intuitive pictorial representations

equivalent to tables were common in the co-ed classes, whereas bar charts were universal in

the girls' class. Of the 66 representations drawn, 26% were Pictographs, 18% were Tables or

Tallies, and 53% were Scaled Graphs. Three percent drew both a Table and a Graph and 9% did not produce a representation. For Code 1 responses, the context was clear but the data were incomplete. For Code 2, all data for the group were recorded in a recognizable structure. As is seen in Figure 7, all representations recorded data as individual measures collected by particular students, rather than combining data in an ordered fashion, which would have been considered as a higher level of transforming data as part of an analysis. Examples shown in Figure 7 show the three types of representation drawn at the two recognisable levels (6% were coded 0).
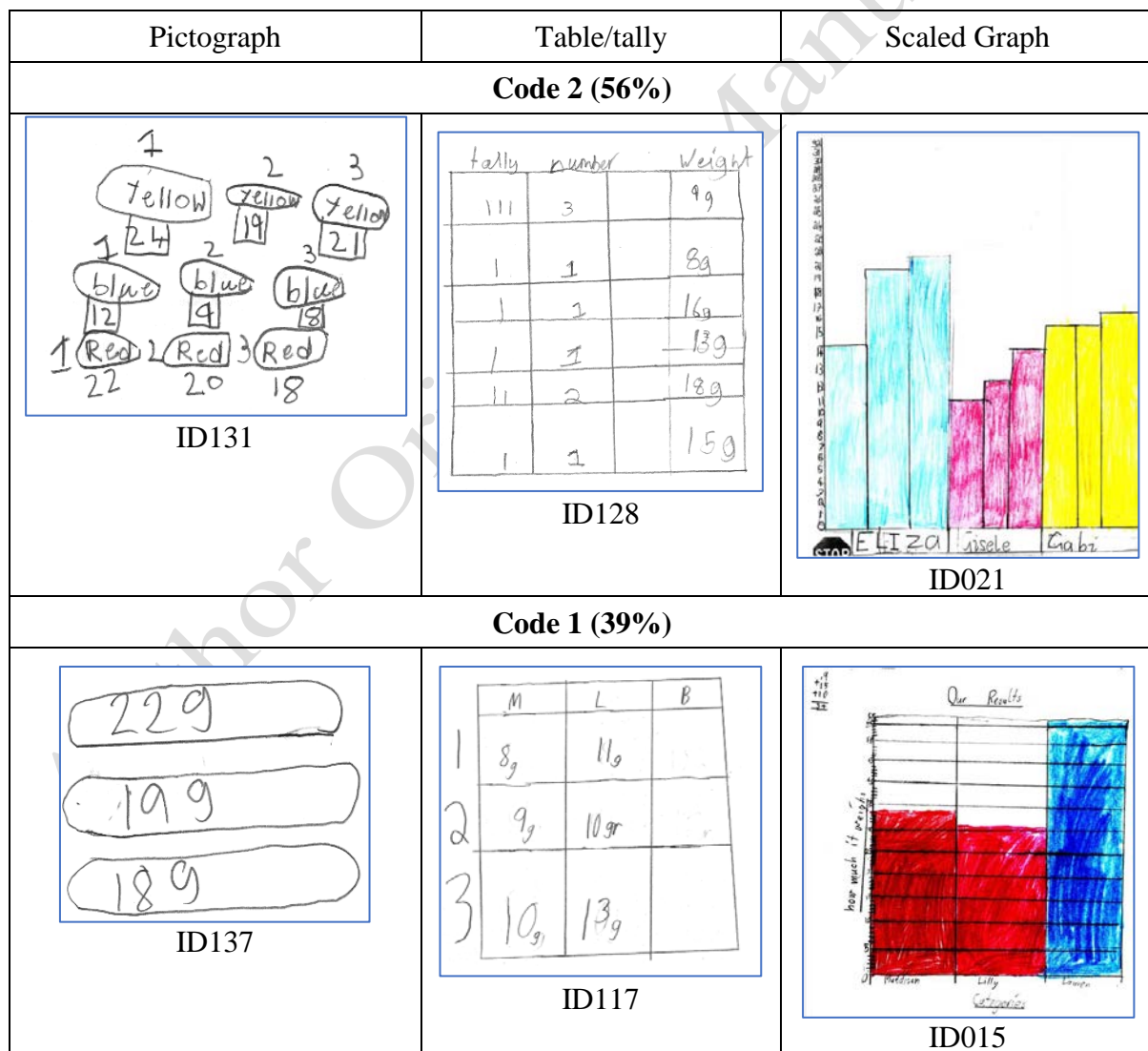
Figure 7. Examples of representations of hand-made licorice stick data, Q5.

When representations were again produced for the machine-made licorice sticks (Q5), the representations were very similar to those produced for the hand-made sticks. These were not influenced by the representation of the hand-made sticks for the class made by the teacher (see Figure 4). Seventy-one percent of representations were of the same type each time, 20% were different, and 9% had one representation missing or one or both uninterpretable. Of the 68 completed representations for the machine-made sticks, 15% were pictographs, 31% were tables or tallies, 50% were graphs, and 4% included a table and picture. Examples for the two main levels are shown in Figure 8 (4% were coded 0).

Pictograph	Table/tally	Graph
Code 2 (81%)		
ID119	ID104	ID007
Code 1 (14%)		
ID144	ID024	ID018
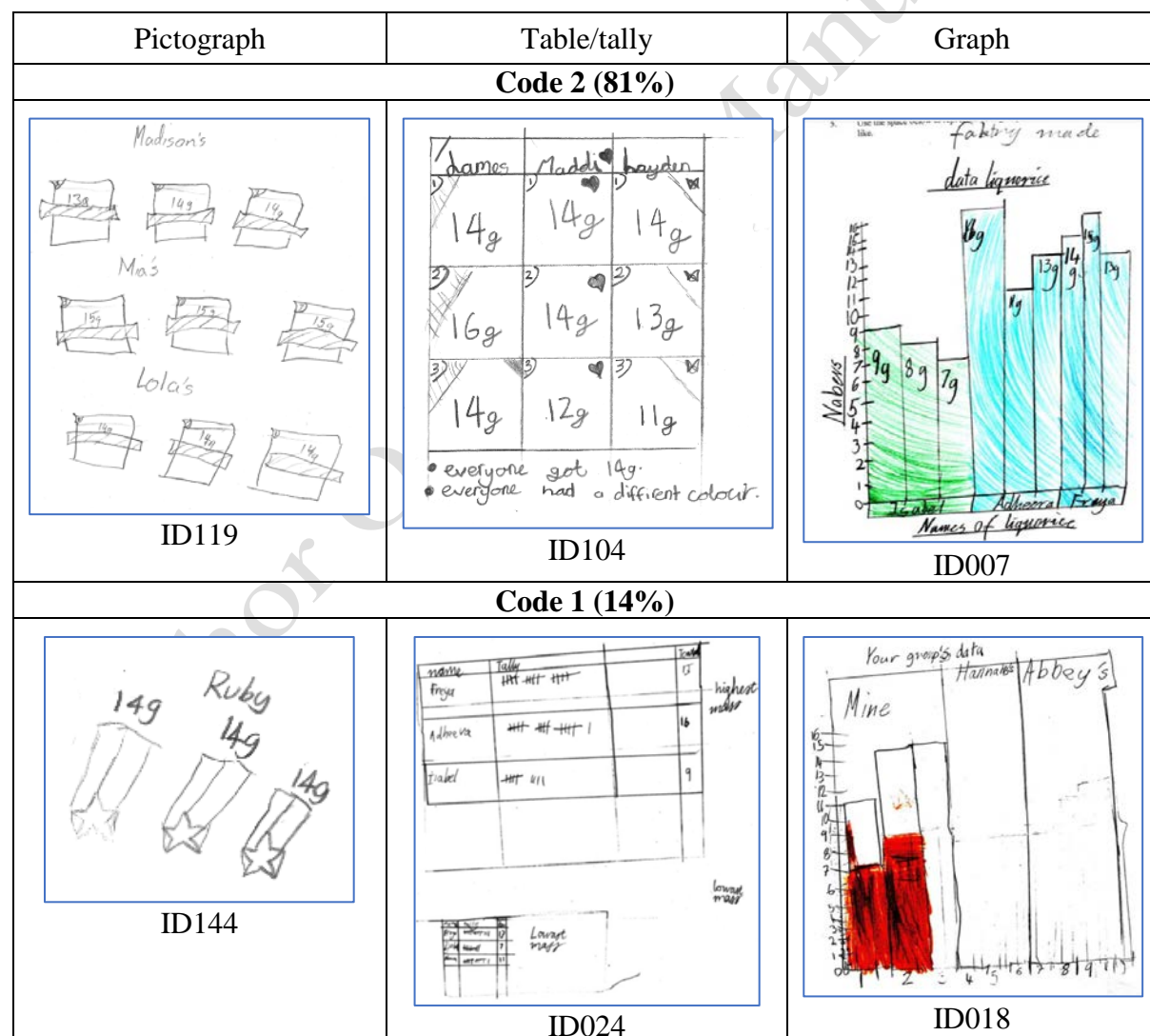
Figure 8. Examples of representations of machine-made licorice stick data, Q5.

25

The next question (Q6) related to the class plots created for the two methods of production and asked for a description of the *shape* of the plot. No other qualification was added to the question. Code 0 responses referred to the plots as "pictures"; Code 1 responses picked a valid element in the plot but ignored any aspect of variation; Code 2 responses linked one element of the plot to the variation displayed; and Code 3 responses combined several elements of the plot reflecting variation. Examples are shown in Table 6 for each method of production and given the question it was not expected that the word "variation" would be used in the responses.

Table 6. *Describing the shape of the class plot, Q6.*

Level	% Hand-made	Examples: Hand-made	% Machine-made	Examples: Machine-made
3	41%	Its got spaces and humps and sections. There is alot at the start of the chart. (ID114) They are mostly around the middle. Our typical number of mass was eleven. The highest number was 30g, the lowest was 5g. (ID016)	56%	It looks like a city because it has small buldings [buildings] and tall buldings [buildings] the highest bulding [building] was 13 grams. (ID001) Rocket, skyscraper, tower, a mega roller. (ID120)
2	17%	There were lots pileing [piling] on top so it looked like mountins [mountains]. (ID009) I think it makes a zig zag. (ID113) Secshons [sections] with spaces. (ID125)	39%	It is nearly a triangle. (ID137) The shape of the class plot is a line with numbers stacked on top of each other and they ended up with humungous piles. (ID015)
1	20%	The class plot has some of the same numbers and it is a start line and in oder [order]. (ID023) Square almost in the middle. (ID121)	1%	2 rectangles stuck together with lots of squares that are agctally [actually] sticky notes! (ID142)
0	21%	Rectangle. (ID140) A square/rectangle or a face. (ID102)	4%	Rectangle. (ID143)

When students were asked more specifically about what the plots told them about the variation in the sticks they had made (Q7), describing this combining the elements provided in the context was considered a Code 2 task, whereas describing a different feature of the data

in the plot not related to variation was considered Code 1. Code 0 responses did not address

the question. Examples are given in Table 7.

Table 7. *Responses explaining what the class plot tells about variation, Q7.*

Level	% Hand-made	Examples: Hand-made	% Machine-made	Examples: Machine-made
2	66%	There are more in the middle near 11 and less nearer to 20 and up. (ID003) It tells me that some licorice sticks weighed more than others. (ID006) More on 10. Less on 7. One less on 13. (ID103)	60%	A lot of people had 13g and 14g. Less people have 10g. (ID110) The most comen [common] is 14 the second most comen [common] is 15, and 16 and 13 are the same. (ID129) Lots and lots of people had between 10g and 16g. (ID124)
1	14%	The class plot has lots of sticky notes pilled [piled] on top of each other because lot's [lots] of people had the same mass. (ID019) That alot of peoples licorice sticks were ten grams. (ID117)	24%	Most peoples weighed 14 grams. (ID145) 13g has the most sticks. (ID102)
0	20%	Numbers and g and note sticker. (ID146) It tells you how many grams the licorice people made was. (ID118)	16%	Star (ID143) With notes with them and the wanght [weight]. (ID146) How hevey thay way. [How heavy they weigh.] (ID148)

The total score possible for Objective 1 questions for both hand-made and machine-

made licorice sticks was 26 (13 for each type). As seen in Figure 9, the median score was 19

with 75% of students achieving a score of 16 or more, with the top 25% doing very well

(with scores of 23 or more).
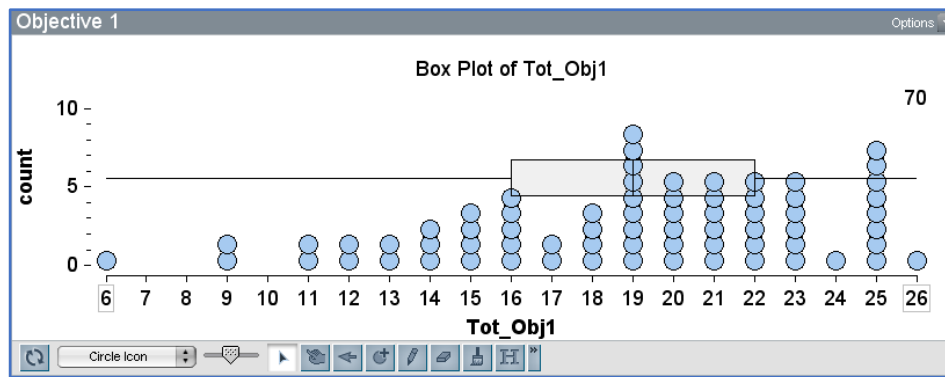
Figure 9. Scores for Objective 1.

In the discussion that followed each way of making licorice, the teacher attempted to consolidate the developing understanding. This is illustrated in the following extracts after machine-made data were represented.

Extract 1

Teacher	[referring to machine-made data]: What can you tell me about the variation? Is there any variation? Is there lots of variation?
S1:	There's not that much variation.
Teacher:	Why do you think there's not that much variation?
S1:	Because it's all between 10 and 17.
Teacher:	But how do we get those sort of numbers, how do we get that sort of data in the first place?
S1:	Because they're all, because they're all the same size, because they come out as one shape?
Teacher:	Ok.
S2:	Because they were machine-made not hand-made because when you have hand-made you can make a big error and smaller….

Extract 2

Teacher:	[referring to the machine-made plot on the wall] For the machine-made one we did today, who can tell me the variance between the first label and the last label?
S3:	Um, well 11, um, to 17 well is 6 grams different from 11, which is the smallest from the biggest so 11 to 17.
Teacher	[referring to the hand-made plot on the wall]: Yep, good … So what would you say about the first ones' variance?
S4:	That they were all different.

Teacher:	Does it look like the one we did today, the first one?
Class Response: No!	
Teacher:	What, what is the difference?
S4:	That there is more, that there is um, that they all spread out and all bunched up in that one.

Extract 3

Teacher:	[referring to the two plots] Do you think that there was more variation using the machine or less?
S5:	Less.
Teacher:	Can you tell us why?
S5:	Because it's a machine like, the machine makes them all about the same size and when you're doing them with your hands you can't really tell if they're going to be the same size or not.

Objective 2: Identifying typical values to make predictions

When asked to describe the typical mass for the class data (Q8), 64% picked a reasonable value for the class plot for hand-made sticks within the range for the data collected in their classes and 77% did so for the machine-made (Code 1). When asked for reasons why this value was typical (Q8 Why), 54% gave meaningful explanations for the hand-made sticks (Code 1), e.g., "Because those two numbers are the most common mass people got" [ID145]. This percentage increased to 61% for the machine-made sticks.

Prediction of the mass if they made another licorice stick by each method (Q9) showed an increased appreciation of the task, with 89% of students making a reasonable prediction within the range for the data collected in their classes for hand-made sticks and 96% doing so for machine-made (Code 1). When asked how they had decided on their predictions (Q9 How), some did not respond, whereas the other codes reflected combining the data element of the task with other increasingly appropriate elements available in making decisions. These ranged from a non-statistical reason (Code 1), to a reason based on the student's own personal data (Code 2),

to a reason incorporating appreciation of the middle of the class data set (Code 3). Examples

and percentages are given in Table 8.

Table 8. *Predictions and responses for how the prediction was made for the mass of one more piece of licorice, Q9.*

Level	% Hand-made	Examples	% Machine-made	Examples
3	27%	14g. Because it is close to the middle. (ID148) 14g. I decided this number because it was in between 10-17. (ID109)	63%	13g/14g. Most people had 13g and 14g. (ID127) 14g. This is a high tower. (ID101)
2	40%	11g. Because the last time I did it I got 11g so I think it would be the same. (ID138) 15g to 19g. Because all my other ones are around fifteen and nineteen. (ID132)	21%	[14g] Because one of my pieces are 13g and 15g, And it would be in between. (ID024) [12g] Because it is in the middle of my two other liquorises [licorices]. (ID022)
1	20%	[8g] I would roll it small. (ID143) [12g] Because it's big and I like getting big bites. (ID135)	9%	[13g] It might be there [their] lucky number. (ID104) [25g] If you love licorice you what [want to] have 30g. (ID146)
0	13%	No response	7%	No response

The total score possible for Objective 2 questions on typical and prediction was 12 (6

for each type of licorice stick). Students found it relatively easy to interpret the class plots in

this way, as is seen in Figure 10, with a median score of 9 and 75% of students scoring 8 or

more.
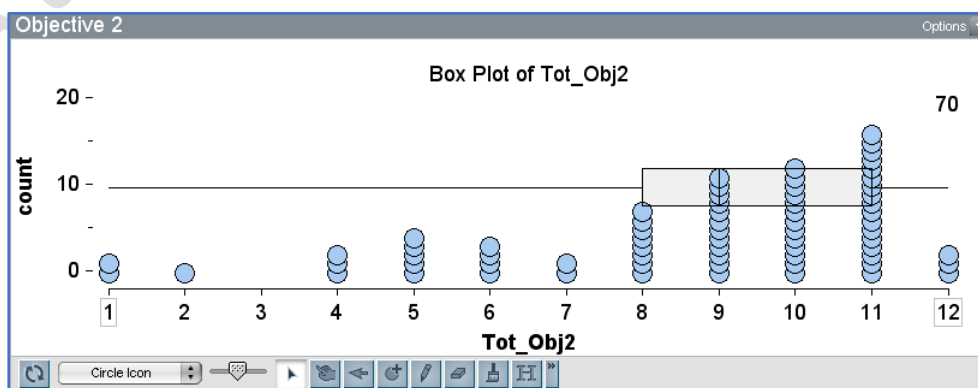
Figure 10. Scores for Objective 2.

Class discussion illustrates the consolidation that took place when students were describing their typical values and moving to making predictions.

Extract 1

Teacher: Girls what do you think would be the next thing for us to do? Everyone has put one [sticky] note on, what could we do next? Have a look at what we've got with those pink [sticky] notes, they're starting to show something.

S6: We could … go and see and count them.

Teacher: You think we could count them. That would make sure that everyone has done one. Do you think we should, have a look at our picture and see what it shows us so far? Is it showing us something? We're looking for this word here; remember we're looking for typical. Does our graph or does our class plot show us something about that?

S7: Um I think it shows that um a certain amount [sic] of people had the certain number, like 4 people had like one number and 2 people had a different number.

Teacher: S2 was just saying that our class plot is showing that 4 people had the same number, and how did we represent that on the class plot? How did we do that?

S7: We put them on top of each other… well it's sort of like making a um, the class' like graph out of the um numbers of the grams of the licorice.

Teacher: Yes, you're right, well done. Anything else you'd like to say S8?

S8: Well the graph shows us that most, the typical number is 11, 11 grams.

Teacher: Yes excellent work. So typical it's starting to show. Okay, so if we, here's a question, you have two more [sticky] notes on your desk, if everyone got a chance to put more [sticky] notes on the plot, do you think it would show something else?

Class: Yes.

Teacher: What might it show?

S9: It would show our entire average because that one only shows part of our average because we've only put one of them, one of our [sticky] notes on but if we put our other two on, then we can actually know our entire average.

Teacher: Excellent work, or our entire typical mass.

31

Extract 2

Teacher: Would you like to tell us what the typical mass was for your group…?

S10:   What do you mean by typical mass?

Teacher: The normal mass for your group.

S10:   So the typical um mass of our licorice sticks would be about 13.

Teacher: Yes, she's saying it's about. Does she have to be exact?

Class Response: No.

Teacher: And why is that important in your answer?

S10:   Um, why is what important?

Teacher: Did everyone get exactly the same in your group?

S10:   No like it's around, like around that number.

Objective 3: Comparing plots to make justifications

The last section of the workbook explored students' appreciation of the difference in the variation between the two types of licorice and their ability to summarise the activity in their own terms. When asked the specific question of whether the two plots were the same, 96% said "no", one student said "sort of," and the others did not respond. When asked to list the differences in the plots, responses ranged from idiosyncratic descriptions unrelated to the purpose of the activity (Code 0), through single non-specific reference to difference without reference to the two data sets (Code 1), one specific comparison, including reference to the two data sets (Code 2), to a comparison of two or more characteristics specifically related to the two data sets (Code 3). Examples and percentages are given in Table 9.

Table 9. *Description of differences between the two plots, Q11.*

Level, %	Examples
3 19%	Tuesday's plot had a lot of veriation [variation]. Thursday's plot had not that much veriation [variation]. Thursday's plot was a lot taller than Tuesday's plot. (ID109) Because 13g had the most with a masheen [machine] and 10 had the most with hand made. They are not the same shape. (ID116)
2 54%	The first ones range was 6 to 20 and the second one was 10 to 16. (ID113) That the first one is all spre[a]d out and the second on[e] is all bunched up. (ID130)
1 14%	One had less veriation [variation]. (ID127) One is higher than the other. (ID101)
0 13%	Because the blue one is hand-made and the pink one is factory-made. (ID005) Some of them are a lot smaller than the other sticky notes. (ID155)

When asked what the difference in the plots told them about two ways of making licorice, some responded with idiosyncratic issues not related to making licorice (Code 0). Some focused on a single feature of the two settings, either difference not specifically related to the making of licorice, or to making licorice but no mention of variation (Code 1). Others noted both the difference of the two methods and the variation or lack of it for the appropriate mechanism (Code 2). Examples and percentages are given in Table 10.

Table 10. *Explaining what the results tell about making licorice sticks, Q12.*

Level, %	Examples
2 56%	It tells me that hand-made wasn't quite as accurite [accurate] as factory-made because the factory makes every stick very much the same and hand-made isn't. (ID006) The factory is easier and is more similar. In weight. Because it is made the same way. (ID016)
1 23%	One of the graphs are tall and the other graph is just ups and downs. (ID114) you kan [can] mack [make] them by hand or by facktre [factory] but thay [they] will be difrent [different]. (ID017)
0 21%	Homemade licorocice [licorice] is more healthy and would be a better choice if you like licorice. Machiene [machine] made licorice is made with lots of sugar and is unhealthy. (ID122) The fun factory was the best because esar (easier) to make. It had beter [better] patens [patterns]. Fun to make. (ID023)

Given the class discussion at the beginning of the activity it was of interest to ask students which type of licorice they would rather buy if the licorice were real (Q13). Thirty-

one percent gave no response, a choice without a reason, or an idiosyncratic choice (Code 0) (e.g., "red because it is the best type of licorice" [ID104]). Thirty-seven percent chose based on a personal preference (Code 1), more commonly hand-made (e.g., "…because it could be really big. I love licorice." [ID101]) but also some machine-made (e.g., "… because they look like the real thing/product." [ID010]). Thirty-one percent gave a statistical response based on lack of variation (Code 2) (e.g., "Machine because if you liked it, all the rest would be the same." [ID143]).

Finally, students were asked to draw a picture showing what they found out in the investigation (Q14 Picture) and to write a story to match the picture (Q14 Story). This was a demanding task for Year 3 students but it was of interest to see how they would cope with it. For these tasks coding included an extension above the Code 3 level (Code 4 in Figure 11 and Table 11). Code 0 responses referred to only one way of making licorice sticks without either a representation (Q14 Picture) or a description (Q14 Story) of the output data (or were blank). Code 1 responses presented only one way of making the sticks including some display or mention of variation. Code 2 responses showed or described two ways of making licorice sticks but there was no further mention about the data and variation. Code 3 responses again referred to both ways of making licorice sticks, with difference visible or described but not specifically as experienced in the activity. Finally, Code 4 responses provided specific information added to the response about the difference in variation between the two methods of making licorice sticks. Examples and percentages of responses are given in Figure 11 and Table 11.
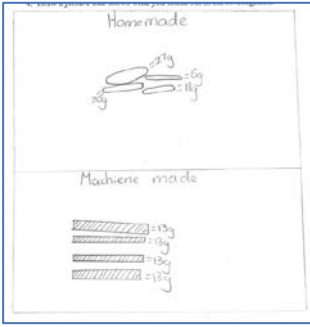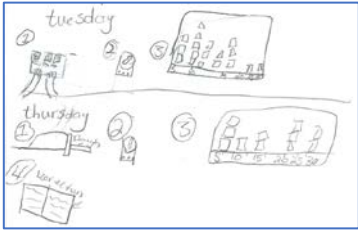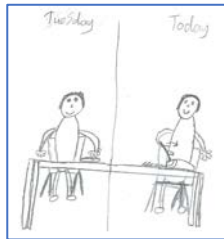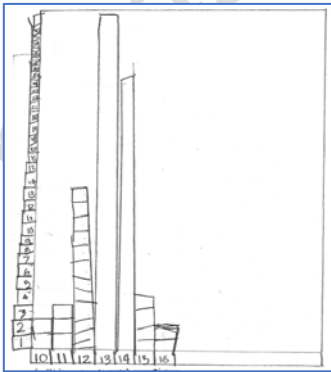
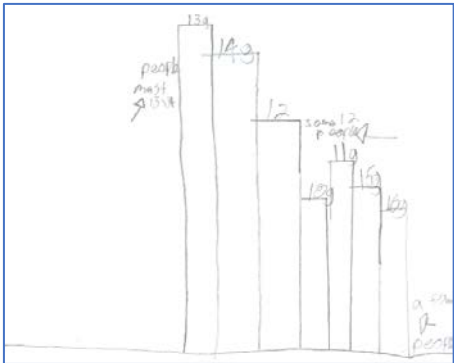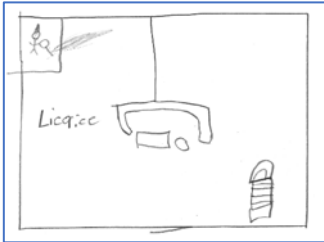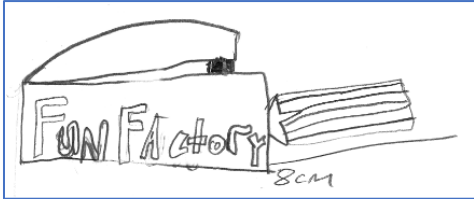| Level, %	Examples
4 21%	 ID122 · ID142
3 14%	 ID130 · ID143
2 16%	 ID119 · ID124
1 20%	 ID104 · ID127
0 29%	 ID109 · ID139

Figure 11. Pictures to show what was found in the investigation, Q14 Picture.

Table 11. *The story to match the picture of the investigation, Q14 Story.*

Level, %	Examples
4 11%	We made licorice by hand on Tuesday and we made licorice on Thursday and I le[a]rnt we get more variation if we make things by hand. (ID12) Machiene [machine] made things are exac[t]ly the same and weigh the same and home-made things are very different and have lots of veriatoin [variation]. (ID122)
3 17%	Factory made liquorice [licorice] is healthyer (healthier) and more average where as hand-made liquorice [licorice] is less healthy and less average. (ID003) One day we made hand made licorice. The nexst [next] 2 days we made machine licorice with a play dough machine and play dough hand made is harder than a machine. (ID123)
2 27%	On Tuesday we made licorice sticks by hand and on T[h]ursday we made licorice sticks by masheen [machine]. (ID116) On Thursday we made licorish [licorice] out of a playdough meshean [machine]. It was made into a very long star. On Tuesday we made licorish [licorice] out of hand into a roll. (ID119) 31 people had 13g, there is a big dif[f]erence between the two graphs. (ID117)
1 19%	1. We made our dough. 2. Then we maked sure it was 8 cm. then we weighed the dough. 3. and then we put our weighing information and stuck it on the wall. (ID130) 14 is the toles [tallest] because the people mite off [might have] put more playdo in the meshen [machine]. (ID131)
0 26%	The machene [machine] is making licorice and the licorice is coming out. (ID109) We found out how to make licorice out of a machine. (ID114)

More than a quarter of students struggled *either* to represent *or* to write a story about the activity; 11% of responses given Code 1 for both tasks; 17% of responses with Code 1 for one task achieved at least a Code 2 response on the other. Further, 44% of responses were at least at the Code 2 level on one of the two parts of the final task.

The total score possible for Objective 3, drawing conclusions from the plots and the two ways of making licorice sticks, was 15. This included drawing a picture and writing a story about their investigations. It was a more difficult task for Year 3 students, and this is seen in the spread in the data in Figure 12. The median is 8, with about 25% scoring 5 or less.
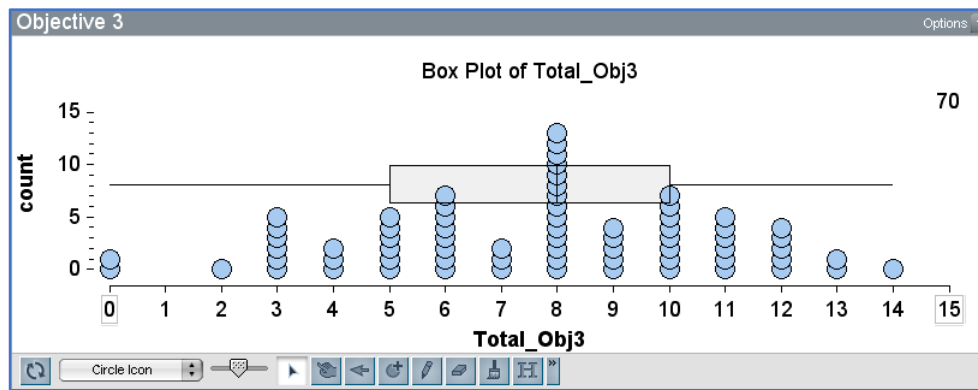
Figure 12. Scores for Objective 3.

The discussion in class illustrates how the differences in the two methods of making licorice sticks were conceived by students at the end of the activity.

Extract 1

S11:   So could I say that the difference was that the range for the first one was 6 to 28, and the range for the second one was 10 to 16?

Researcher: That's definitely one of the differences, yep, good job.

Extract 2

S12:   … can I say something like, um, does it mean like saying machine made are more controlled? Is that what it basically means? And hand-made was… um… had a lot of difference.

Researcher: Was what?

S12:   Hand-made had lots of differences.

Researcher: Yes, very good.

Extract 3

Teacher: So, are the two plots the same? Yes or no, Year 3?

Class: No.

Teacher: Can you tell us why or why not or how they are different?

S13:   Um the blue plot it's sort of stacked up on top of each other like a storeroom, except it hasn't got a lot of space and the hand-made plot, it's sort of, it's same again like a storeroom except it's got tons of space so it's more spread out.

Extract 4

> Teacher: … would anyone else like to give us an answer on how these two plots are different?
>
> S14: They're quite different because in the factory-made they're quite close together the results and in hand-made they're spread out.
>
> Teacher: Yes, they're spread out, why were they spread out, can you give a reason for your answer?
>
> S14: It's probably because hand-made is not as accurate as factory-made so people were getting all different answers… that is why hand-made is more, is more spread out.

The result of combining the scores for the workbook entries for the three objectives is shown in Figure 13. With a total possible score of 53, 88% of students scored over half and 50% scored 37 or better. Although as can be seen in Figure 12, two students did not score on Objective 3, Figure 13 indicates better performance by those students across the other two objectives.
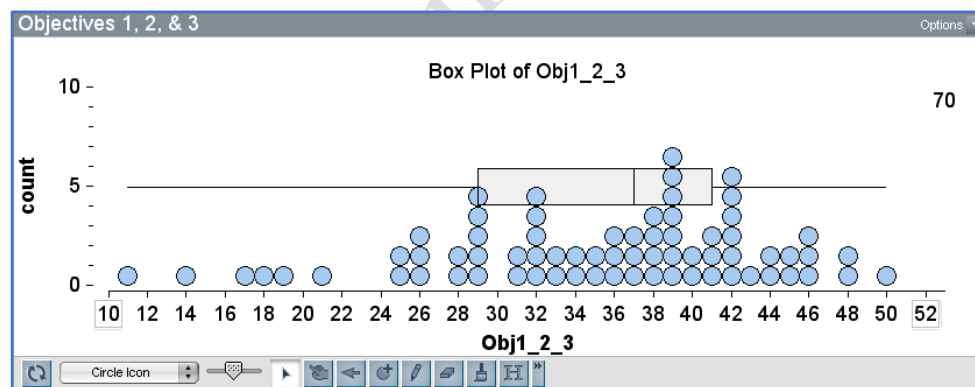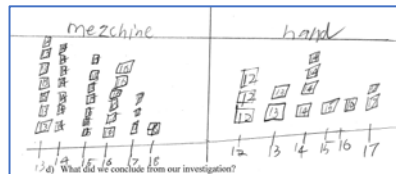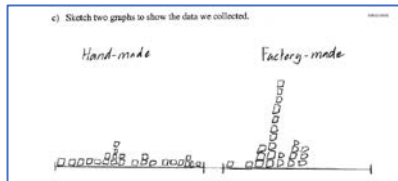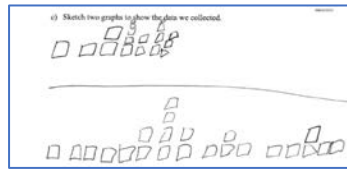

Figure 13. Sum of scores for Objectives 1, 2, and 3.

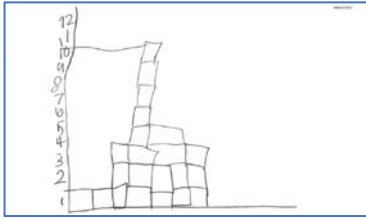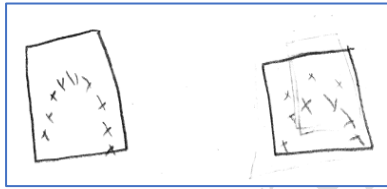## Objective 4: Retaining concepts introduced in the licorice activity

Students' retention of understanding of the activity was gauged from the end-of-year survey of all students and interviews with the students in the girls' school.

End-of-year survey. The four questions on the end-of-year survey (see Table 2) were intended to assess retention of student understanding in relation to the objectives of the research four months later. The explanations and drawings were not as extensive as those

given at the time of the activity and the coding reflected recognition of a singular (Code 1) or multiple (Code 2) use of the elements from the task, or further, whether they were related together meaningfully (Code 3), for example labelling plots. Table 12 shows examples of each level, with percentages.

Table 12. *Responses to the end-of-year survey (n=67).*

Level, %	Examples
Q1: Explain how we collected the data for the activity.	
3 6%	We made factory made and hand made and wighed [weighed] it. Then we recored [recorded] it on posted [sticky] notes. (ID009) We made handmade licorice that had alot [a lot] of variation and machine made that was pretty much the same. (ID152)
2 50%	We weighed the licorice on a scale to see the weight of the licorice. (ID002) We wrote down how much our licorice weighed and put the measurments [measurements] into a number line. (ID006)
1 20%	We made them and mesherd [measured] them. (ID010) We meshered [measured] the lenth [length] and the with [width]. (ID121)
0 24%	From a plot. (ID004) We used sticky notes to put on a graf [graph]. (ID136)
Q2: What data did we collect/and record?	
2 8%	The highest weight for the hand-made is 30g and the highest weight for factory-made is 16g. (ID002) That the tipacal [typical] mass was fiffteeng [fifteen] for factory made and 11g for hand made. (ID008)
1 68%	How heavy it is. (ID149) We collected the weight of our licorice and made a graph on the board. (ID016)
0 24%	We worked in groups. (ID103) [We] put up a very big piece of paper and each person had a turn. (ID147)
Q3: Sketch two graphs to show the data we collected.	
3 11%	ID128 ID141
2 32%	ID001 ID022

1 24%	 ID017	 ID123
0 33%	 ID146	 ID131
Q4: What did we conclude from our investigation?		
3 23%	The factory made liquorice [licorice] sticks had less variation than hand-made ones. (ID015) They were different because the Hand-made was spread out than the other and the factory-made is taller than the other. (ID001)	
2 9%	That factory made is more squished together and hand made is spread out. (ID002) That the factory made is more eficent [efficient] than the hand made. (ID012)	
1 24%	That most people got 13. (ID104) We did hand made licorice and michene [machine] made licorice and I reckon hand made was easier. (ID109)	
0 44%	We found the data on the graph. (ID126) There were alot [a lot] of both factory made and hand made. (ID022)	

The total score possible for the end-of-year survey questions about the Licorice activity was 11 but the highest score was only 6, indicating that the students had extra difficulty in the context of a survey (see Figure 14), which also included 18 questions on other topics related to statistical literacy. The students could visualise the activity and usually give at least one specific descriptor of the activity, which is expected cognitively by Year 3 [48, p.25]. The correlation between the total of the scores for Objectives 1, 2, and 3, with the total for Objective 4 was 0.486 ($p < 0.0001$), indicating a moderate relationship between responses during the activity and the end-of-year survey responses.
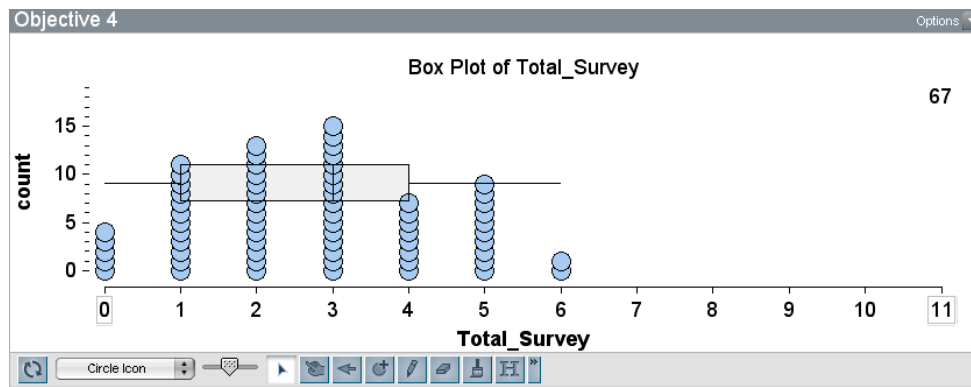
Figure 14. Scores for Objective 4.

End-of-year interviews. Because variation was the critical statistical concept being introduced in the Licorice activity, the interviews at the girls' school focussed on it while revisiting the outcomes of the complete activity. Examples of responses for each category described in Table 3 are given in Table 13. The percentages add to more than 100% because 10 students (45%) gave responses in both Categories A and B, whereas 2 students (9%) gave responses in both A and C. For the question on what they learned from the activity, examples in the four categories described in Table 3 are also given in Table 13. Only 4 students (18%) gave responses in two of the four categories, none including Category G.

Table 13. *Responses to end-of-year interviews (n=22).*

Code, %	Examples
Based on plot for hand-made sticks (Figure 2): Describe the shape	
A 86%	… like a city … lots of tall ones and some small ones. (ID018) It's like something going down up, down up, like cities, houses, cars, and people. (ID023)
B 45%	Like a wriggly line … cause its all spread out. (ID005) Like most of them … are like separated a lot. (ID013)
C 9%	The biggest area is actually around the 10 grams. (ID004) Most of them did I think 10 grams or maybe 13. (ID007)
Explicit understanding of Variation	
Yes 64%	It goes up to around 30 grams and the low was 5 grams … quite a bit of variation. (ID002)
No 36%	
What did you learn?	
D 14%	Didn't really think about what I learnt cause I had so much fun. (ID006)
E 23%	I learnt some new words … . (ID009) I learnt words. (ID001) Graphs and everything. (ID017)
F 27%	How to weigh stuff and just to … like measure it as in 1 cm … . (ID008) Well, I learnt about maybe masses. (ID014) I learnt that the engineering and that engineers have to weigh, try, and get the same licorice weights. (ID023)
G 55%	There can be all different kinds of variation. (ID004)? It's hard to predict … it might not even be just one exact spot … (ID007) I learnt that not everything is the exact same size … (ID011)

Discussion

The purpose of using the Manufacturing Licorice activity as the focus of this study was multi-faceted. Using a STEM-related context was intended as a starting point for creating an awareness of STEM contexts through knowledge of the technology/engineering aspects of the "manufacturing," seeing uses for students' emerging mathematical skills in measuring length and mass, and developing careful habits of measurement important for both science and engineering. Students experienced a data modelling procedure without it being made explicit, but with a major focus on variation. Through looking at variation they began to

focus on expectation in the form of typical values in the class representations of the data. This focus first on variation reflects both the fundamental claim of Cobb and Moore [19] that statistics exists because of variability and the research of Watson and colleagues [e.g. 25–26] confirming that children develop appreciation for variation naturally before appreciation of expectation. Typically in school curricula related to statistics this has not been the order of introduction across the school years [20]. The interviews carried out with the students in the girls' school demonstrate this natural tendency related to variation first when asked to describe the shape of the plots in Figures 4 and 6, with many more girls mentioning the variation in frequencies or spread of the data, than mentioning central expectation.

The general research question was related to the capabilities of the Year 3 students in the study to demonstrate understanding of the fundamental statistical concept of variation through data modelling in a STEM-related context. To answer this question, the four specific objectives of the research were related to describing and representing the variation experienced, suggesting typical and predicted values from the data collected, reporting on the difference in variation within the two methods, and reporting on the activity four months later. In relation to the use of the SOLO framework [48], the coding of responses mainly reflected Unistructural or Multistructural understanding with a few Relational opportunities. This reflects the expectations for the beginning understanding of young students entering the Concrete Symbolic mode of development. Figures 9, 10, and 12 demonstrate that most students displayed quite reasonable performances on the first three objectives. Although the correlation between the sum of scores for the first three objectives and the fourth was highly significant, the level of end-of-year retention of the basic capabilities shown during the activity was somewhat disappointing (cf. Figure 13). Several factors should be kept in mind with regard to this result. The students had not previously experienced completing a survey of this type. Within six A-4 pages, there were nine questions, made up of 22 parts, including the

four on the Manufacturing Licorice activity. Five were short answer or multiple choice, but the rest asked for some explanation. The teachers did not impose a time limit and some students were given more than one opportunity to complete the survey. The literacy levels of some of the students, who were at the end of Year 3, however, made it difficult for them to express their ideas in writing. Although the objectives of the Manufacturing Licorice activity were ambitious, the feedback from the teachers was overwhelmingly positive, with comments made about "variation" being mentioned later in other classroom activities.

The Manufacturing Licorice activity was innovative because it provided a context where the targeted expectation or typical value for both the hand-made and machine-made products was the same. This scenario was set up within the activity where students were required to make licorice with the same dimensions by both methods. Previous studies have usually focused on difference in expectation when comparing two groups of data [e.g. 49– 51]. Often the contexts have been related to comparing attributes of girls and boys or different "treatments." These types of contexts provide situations where the expectation for each group is different. For example, the number of meaningful words memorised would be expected to be greater than the number of nonsense words memorised [52] or the reaction time of Year 10 boys would be expected to be quicker than that of Year 5 boys [53]. Often in these situations students focus on the mean or median difference, with little consideration of the variation in the two data sets. In this study having the expectation that the licorice sticks made in either manner would have similar masses removed the complication of comparing "middles" and allowed students to focus on identifying spread, and hence variation.

As well as introducing a meaningful context as a basis for STEM education [e.g. 4–5], the activity in this research also satisfied many of the discipline content descriptions for Year 3 in the country of origin. For example, in *The Australian Curriculum* [45] in the Mathematics section, there are descriptors related to measurement including length and mass,

to identifying data sources and planning data collection and recording, and to interpreting and comparing data displays. In Science, Science as a Human Endeavour involves making predictions and describing patterns and relationships. Further, Science Inquiry Skills suggest many of the skills that apply to the investigation the students undertook: consider the elements of fair tests and use formal measurements to make and record observations accurately; use a range of methods including tables and simple column graphs to represent data and to identify patterns and trends; and compare results with predictions, suggesting possible reasons for findings. With respect to the Design and Technologies Curriculum, one of the two strands of the curriculum encompasses the foundations required for Technology and Engineering: knowledge and understanding: "the use, development and impact of technologies and design ideas across a range of technologies contexts." This knowledge and understanding begins in Year 3 with the kinds of hands-on experiences and class discussions that were provided in this project.

Besides these content descriptors across the *Australian Curriculum*, from the Mathematics curriculum, students should gain experience in the proficiency of *Reasoning*, "when they explain their thinking, when they deduce and justify strategies and conclusions reached, when they adapt the known to the unknown, when they transfer learning from one context to another, … and when they compare and contrast related ideas and explain their choices" [45]. Other countries have similar goals in their curricula in subjects related to STEM. In the United States, for example, the *Next Generation Science Standards* [16] addresses reasoning many times in its Science and Engineering Practices and the *Common Core State Standards for Mathematics* [54,p.6] advocate Mathematical Practices based on "problem solving, reasoning, … communication, representation, and connections." All of these apply to context-based STEM activities and were exhibited to some degree by Year 3 students in this study. As well, the research incorporated two of the Guiding Principles of the

Early Childhood STEM Working Group [18] in focussing on the selection of a context in which students were inherently curious and on the children's developing capabilities in representation and communication.

Conclusion

The general research question and its specific objectives related to the Manufacturing Licorice activity were met in terms of documenting Year 3 students' capabilities to engage with a meaningful context related to STEM—a context meaningful for Year 3 that exhibited variation and provided a foundation for developing the statistical modelling skills essential for more advanced investigations. The fact that well over three-quarters of the students could achieve over half of the total score across the three main objectives assessed in their workbook responses is encouraging. Although the written responses on the end-of-year survey were somewhat disappointing, the individual interviews with the students in the girls' school revealed that they could easily discuss the context of the activity and respond well to prompts from the interviewer in relation to describing ideas related to variation and typicality. As one of the first studies linking these fundamental ideas in relation to a STEM context, it provides benchmarks for future research with young children. The next stage is research on students' developing understanding as they experience more complex environments involving STEM disciplines and further consolidate their statistical understanding.

Acknowledgements

References

1. Office of the Chief Scientist. Science, technology, engineering and mathematics in the national interest: a strategic approach. Canberra (AU): Australian Government; 2013. Available from: http://www.chiefscientist.gov.au/wp-content/uploads/STEMstrategy290713FINALweb.pdf

2. Office of the Chief Scientist. Science, technology, engineering and mathematics: Australia's future. Canberra (AU): Australian Government; 2014. Available from: http://www.chiefscientist.gov.au/wp-content/uploads/STEM_AustraliasFuture_Sept2014_Web.pdf

3. Pan American Development Foundation. STEM in the Americas. Washington (DC): Author; 2012. Available from: https://www.padf.org/stem/

4. Education Council. National STEM school education strategy, 2016-2026: a comprehensive plan for science, technology, engineering and mathematics education in Australia. Carlton (AU): Author; 2015.

5. Honey M, Pearson G, Schweingruber H, editors. STEM integration in K-12 education: status, prospects, and an agenda for research. Washington (DC): National Academy of Sciences; 2014. Available from: https://www.nap.edu/catalog/18612/

6. English LD. Advancing elementary and middle school STEM education. Int J Sc & Math Educ. 2017. On-line first: DOI10.1007/s10763-017-9802-x

7. Fitzallen N. STEM Education: What does mathematics have to offer? In: Marshman, M, Geiger, V, Bennison A, editors. Mathematics education in the margins. Proceedings of the 38th annual conference of the Mathematics Education Research Group of Australasia; Sunshine Coast. Sydney:MERGA; 2015. p.237-244.

8. Treacy P, O'Donoghue J. Authentic integration: A model for integrating mathematics and science in the classroom.Int. J. Math Educ Sc & Tech. 2014;45(5):703-718.

9. Becker K, Park K. Effects of integrative approaches among science, technology, engineering, and mathematics (STEM) subjects on students' learning: a preliminary meta-analysis. J STEM Educ. 2011;12(5 & 6):23-37.

10. English LD. STEM education K-12: perspectives on integration. Inc J STEM Educ. 2016;3(3). Available from https://doi.org/10.1186/s40594-016-0036-1

11. Kelley TR, Knowles JG. A conceptual framework for integrated STEM education. Int J STEM Educ. 2016;3(11). doi: 10.1186/s40594-016-0046-z

12. Stohlmann M, Moore TJ, Roehrig GH. Considerations of teaching integrated STEM education. J Pre-Coll Eng Educ Res. 2012;2(1):28-34. doi: 10.5703/1288284314653

13. Wang HH, Moore TJ, Roehrig GH, Park MS. (2011). STEM integration: the impact of professional development on teacher perception and practice. J Pre-Coll Eng Educ Res. 2011;1:1-13.

14. Shaughnessy JM. STEM: an advocacy position, not a content area. Reston (VA): National Council of Teachers of Mathematics. 2012. Available from https://www.nctm.org/News-and-Calendar/Messages-from-the-President/Archive/J_-Michael-Shaughnessy/STEM_-An-Advocacy-Position,-Not-a-Content-Area/

15. Watson JM. Linking science and statistics: curriculum expectations in three countries. Int J Sci Math Educ. 2017;15(6):1057–1073.

16. National Research Council. Next generation science standards: for states, by states. Washington (DC): The National Academies Press; 2013.

17. Rao CR. Teaching of statistics at the secondary level: an interdisciplinary approach. Int J Math Educ Sci Technol. 1975;6:151–162.

18. Early Childhood STEM Working Group. Early STEM matters: providing high-quality STEM experiences for all young learners. Chicago (IL): UChicago STEM Education and Erikson Institute; 2017.

19. Cobb GW, Moore DS. Mathematics, statistics, and teaching. Am Math Mon. 1997;104:801–823.

20. Shaughnessy JM. Missed opportunities in research on the teaching and learning of data and chance. In: Biddulph F, Carr K, editors. People in mathematics education. Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia. Vol. 1; Waikato (NZ): MERGA; 1997. p. 6–22.

21. Lehrer R, Schauble L. Modeling natural variation through distribution. Am Educ Res J. 2004;41(3):635–680.

22. Shaughnessy JM, Pfannkuch M. How faithful is Old Faithful? Statistical thinking: a story of variation and prediction. Math Teach. 2002;95:252–259.

23. Watson JM, Kelly BA. The winds are variable: student intuitions about variation. Sch Sci Math. 2005;105:252–269.

24. Watson JM, Callingham RA, Kelly BA. Students' appreciation of expectation and variation as a foundation for statistical understanding. Math Think Learn. 2007;9:83–130.

25. Watson JM, Kelly BA. Expectation versus variation: Students' decision making in a chance environment. Can J Sci Math Technol Educ. 2004;4:371–396.

26. Watson JM, Kelly BA. Can grade 3 students learn about variation? In: Phillips B, editor. Developing a statistically literate society? Proceedings of the 6th International Conference on the Teaching of Statistics, Cape Town. Voorburg (The Netherlands): International Statistical Institute; 2002. Available from: http://iase-web.org/documents/papers/icots6/2a1_wats.pdf

27. Watson JM. Variation and expectation as foundations for the chance and data curriculum. In: Clarkson P, Downton A, Gronn D, et al., editors. Building connections: theory, research and practice. Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia; Melbourne. Sydney: MERGA; 2005. p. 35–42.

28. Watson JM, Callingham RA. Statistical literacy: a complex hierarchical construct. Stat Educ Res J. 2003;2(2):3–46. Available from: http://iase-web.org/documents/SERJ/SERJ2(2)_Watson_Callingham.pdf

29. Watson J, Fitzallen N, Fielding-Wells J, et al.. The practice of statistics. In: Ben-Zvi D, Makar K, Garfield J, editors. International handbook of research in statistics education. Cham (Switzerland): Springer; 2018. p. 105-137. doi.org/10.1007/978-3-319-66195-7

30. Makar K, Rubin A. A framework for thinking about informal statistical inference. Stat Educ Res J. 2009;8(1):82–105. Available from: http://iase-web.org/documents/SERJ/SERJ8(1)_Makar_Rubin.pdf

31. English L, Watson J. Exploring variation in measurement as a foundation for statistical thinking in the elementary school. Int J STEM Educ. 2015;2(3). DOI: 10.1186/s40594-015-0016-x

32. Konold C, Miller CD. TinkerPlots: dynamic data exploration [Computer software, Version 2.2]. Emeryville (CA): Key Curriculum Press; 2011.

33. Watson J, English L. Expectation and variation with a virtual die. Aust Math Teach. 2015;71(3):3–9.

34. Lehrer R, Romberg T. Exploring children's data modeling. Cogn Instr. 1996;14(1):69–108.

35. Crites T, St Laurent R. Putting essential understanding of statistics into practice in grades 9-12. Reston (VA): National Council of Teachers of Mathematics; 2015

36. English LD. Young children's early modelling with data. Math Educ Res J. 2010;22(2):24–47.

37. Gavin, J. How much traffic? Beep! Beep! Get that car off the numberline! In: R. Lehrer R, Schauble L, editors. Investigating real data in the classroom: Expanding children's understanding of mathematics and science. New York (NY): Teachers College Press; 2002. p.39-47.

38. Konold C, Finzer W, Kreetong K, et al. Modeling as a core component of structuring data. Paper presented at the annual meeting of the Research Conference of the National Council of Teachers of Mathematics; New Orleans (LA). April, 2014.

39. Petrosino AJ, Lehrer R., Schauble L. Structuring error and experimental variation as distribution in the fourth grade. Math Think Learn. 2003;5(2&3):131-156.

40. Lehrer R, English L. Introducing children to modeling variability. In: Ben-Zvi D, Makar K, Garfield J, editors. International handbook of research in statistics education. Cham (Switzerland): Springer; 2018. p. 229–260. doi.org/10.1007/978-3-319-66195-7

41. Franklin C, Kader G, Mewborn D, et al. Guidelines for assessment and instruction in statistics education (GAISE) report: a preK-12 curriculum framework. Alexandria (VA): American Statistical Association; 2007. Available from: http://www.amstat.org/education/gaise/

42. Creswell JW. Research design: quantitative, qualitative and mixed method approaches. 2nd ed. Thousand Oaks (CA): SAGE Publications; 2013.

43. Watson J, Skalicky J, Fitzallen N, et al. Licorice production and manufacturing: all-sorts of practical applications for statistics. Aust Prim Math Class. 2009;14(3):4–13.

44. Australian Curriculum Assessment and Reporting Authority. The Australian Curriculum. Version 8.3. Sydney (AU): Author; 2017. Available from: https://www.australiancurriculum.edu.au/

45. Konold C, Harradine A. Contexts for highlighting signal and noise. In: Bender P, Hochmuth R, Fischer PR, et al., editors. Using tools for learning mathematics and statistics. Heidelberg: Springer Spektrum; 2014. p. 237–250.

46. Konold C, Lehrer R. Technology and mathematics education: an essay in honor of Jim Kaput. In: English L, editor. Handbook of international research in mathematics education. 2nd ed. New York (NY): Routledge; 2008. p. 49–72.

47. English LD. Manufacturing licorice: modeling with data in third grade. In: Galindo E, Newton J, editors. Synergy at the crossroads: future directions for theory, research, and practice. Proceedings of the 39th annual meeting of the North American Chapter of the International Group for the Pscyhology of Mathematics Education; 2017 Oct 5–8; Indianapolis (IN): Hoosier Association of Mathematics Teacher Educators; 2017. p. 1040–1047)

48. Biggs JB, Collis KF. Evaluating the quality of learning: the SOLO taxonomy. New York (NY): Academic Press; 1982.

49. Pfannkuch M. The role of context in developing informal statistical inferential reasoning: a classroom study. Math Think Learn. 2011;13,27–46.

50. Rossman AJ. Reasoning about informal statistical inference: one statistician's view. Stat Educ Res J. 2008;7(2):5–19. Available from: http://iase-web.org/documents/SERJ/SERJ7(2)_Rossman.pdf

51. Watson JM, Moritz JB. The beginning of statistical inference: comparing two data sets. Educ Stud Math. 1999;37:145–168.

52. Shaughnessy JM, Chance B, Kranendonk, H. Focus on high school mathematics: reasoning and sense making in statistics and probability. Reston (VA): National Council of Teachers of Mathematics; 2009.

53. Watson JM. Resampling with TinkerPlots. Teach Stat. 2013;35(1):32–36.

54. Common Core State Standards Initiative. Common core state standards for mathematics. Washington (DC): National Governors Association for Best Practices and the Council of Chief

State School Officers; 2010. Available from:
http://www.corestandards.org/assets/CCSSI_Math%20Standards.pdf