Mathematics Education Research Journal (2023) 35:961–999
https://doi.org/10.1007/s13394-022-00421-1

ORIGINAL ARTICLE

Consolidating understanding of variation as part of STEM: experimenting with plant growth

Jane Watson[1,2] · **Suzie Wright**[1] · **Noleine Fitzallen**[1] · **Ben Kelly**[1]

Received: 19 August 2021 / Revised: 18 May 2022 / Accepted: 23 May 2022 /
Published online: 4 July 2022

Abstract

Integrated STEM activities are espoused as appropriate for enhancing student learning in relation to statistical concepts; however, a greater understanding of the way in which students' ideas about those concepts develop is needed to maximise the learning potential offered by engagement in STEM activities. For this study, plant growth was chosen as a topic from the Year 6 Australian Science Curriculum as an appropriate context to employ aspects of the four STEM disciplines to explore students' developing ideas about variation. Sixty-four Year 6 students across three school terms worked in groups of four to trial various treatments and their effects on the growth of radish or wheat seeds. This report considers two aspects of student learning related to this topic based on (i) the formative assessment of features of students' workbook entries specifically related to variation during the part of the classroom activity based on their *TinkerPlots* graphs and (ii) the later summative evidence of learning in responses to end-of-year questions on the activity for 56 of the students. The workbook entries are presented via a qualitative analysis to provide evidence of the forming of understanding of variation in a STEM context, with the SOLO Taxonomy being employed to assess the longer-term evidence and developmental nature of that learning. Overall, a broader picture has emerged of the potential for developing appreciation of variation in a STEM context in primary school.

Keywords STEM education · Variation · SOLO · Plant growth · *TinkerPlots* · Practice of statistics

✉ Jane Watson
jane.watson@utas.edu.au

Noleine Fitzallen
noleine.fitzallen@utas.edu.au

Ben Kelly
ben.kelly@utas.edu.au

[1] University of Tasmania, Hobart, Australia

[2] School of EducationlCollege of Arts, Law and Education, University of Tasmania, Private Bag 66, Hobart, TAS 7001, Australia

Introduction

Much has been written recently about the importance of the STEM disciplines for advances in many fields, from those aiding economic progress in a rapidly changing technological world to those enhancing sustainability outcomes in the face of climate change (e.g. Engler, 2012; Office of the Chief Scientist, 2013). Such moves have fostered the spread of interest in STEM as an integrating force for Science, Technology, Engineering, and Mathematics, throughout the school curriculum, as well as in the industries themselves (Honey et al., 2014). From early childhood (Early Childhood STEM Working Group, 2017), to primary school (Mildenhall et al., 2019; Tytler et al., 2021), middle school (Pecen et al., 2012; Stohlmann et al., 2012), and high school (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2016; Harland, 2011), resources have become available to assist teachers and students to appreciate the potential linkages among these disciplines and the supports that exist to aid investigations in authentic contexts. Indeed, the importance of integration across the STEM disciplines is the focus of Anderson and Li's (2020) collection of reports of various approaches to integration, how the approaches are designed for students (e.g. Wang et al., 2020), and the implications for teacher education (e.g. Delen et al., 2020). Input from around the world illustrates the impact of moves to STEM integration internationally (e.g. Steffensen, 2020).

In this context, Watson et al. (2020a) have made the case for the critical role of statistics in integrating STEM education. The first evidence put forward for this case is based on variation being an essential part of STEM, underpinning the implementation, quantifying, and improving of projects relying on measurement of data in these fields. Through graphical representations and numerical statistical calculations, variation is analysed and decisions made to improve STEM initiatives. Specifically, variation underpins the "practice of statistics" (Watson et al., 2018), for example as outlined by the GAISE report (Bargagliotti et al., 2020):

- Formulate questions, anticipating variability and making predictions,
- Collect data, planning for variability and fair testing,
- Analyse data, accounting for variability using a range of representations and technologies,
- Interpret and evaluate results, allowing for variation and suggesting improvements.

This process, however, is parallel to the Engineering Design Process (Lyden et al., 2018) and Science Inquiry Skills (ACARA, 2019). This natural alliance is further support for considering interdisciplinary opportunities across these disciplines (e.g. Mayes, 2019).

More research, however, is needed to illustrate the way in which students engage with the learning experiences and to provide evidence of the way in which student learning about variation is fostered within multi curricula/discipline learning contexts (Tytler et al., 2019). This is particularly true in relation to the

Australian Curriculum: Mathematics (ACARA, 2019). Although the Proficiencies across the Years 1–6 related to statistics (see Appendix 1) generally cover the fluency, reasoning, and problem-solving required of the "practice of statistics", the Content Descriptions often do not supply the detailed techniques required to carry out investigations in other subject areas. For example, following the mention of column graphs, dot plots, and tables in Year 5 (ACMSP119), in Year 6, this is extended to "Interpret and compare a range of data displays, including side-by-side column graphs for two categorical variables" (ACMSP147). Although using digital technologies is mentioned, it is in relation to those types of representation listed. In this study, it was necessary to extend students' experiences beyond the specifics of the curriculum. In doing this, the various manifestations of variation as encountered in an actual scientific experiment were the focus of the workbook questions as the students completed an activity on Plant Growth. Furthermore, the assessment of the learning that took place at the end of the year employed more general questions, allowing students opportunities to display their appreciation of the impact of variation on the outcomes of the experiments. Background for these two perspectives is considered next.

Variation

The word variation itself has different, or varying, interpretations. In the *Chambers Dictionary* (Kirkpatrick, 1983, p. 1437), variation is defined under the word "variable", in part as, "a varying; a change; continuous change; … departure from the mean or usual character; the extent to which a thing varies; … an inequality in the moon's motion discovered by Tycho Brahe; a change in the elements of an orbit by the disturbing force of another body." James and James (1959) with reference to statistics, equate variation with dispersion and describe it as "scatteration of the data; the lack of tendency to concentrate or congregate" (p. 125). Many sources ignore the term and concentrate on related words like "variability" or "variable". Some texts accept variation as an undefined term. Kitchens (1998), for example, defines a variable as "any characteristic that can be measured on each experimental unit in a statistical study" (p. 3), then claims "It is the *variation* exhibited by the variable that is of interest to the statistician", and further defines, "The *distribution* of a variable specifies the distinct values that the variable assumes and how often these values occur. The distribution illustrates the pattern of the variation in the data" (p. 9). When introducing contextual data to children in primary school, the idea of variation as change or scatter and the need to illustrate (or represent) this in a distribution to study and explain it is the critical starting point in answering questions.

Although statistics is built on variation, at times in the early days of introducing statistics at the school level, it was taken for granted (e.g. Holmes, 1980). Green (1993), and particularly Shaughnessy (1997), brought it to the fore, asking for research specifically on the topic. Shaughnessy (2007) summarised eight conceptions of variability arising from research at that time, of which six were considered relevant to the experiences of the students in the study reported here: variability (i) in particular values, (ii) as change over time, (iii) as a whole range of possible

values, (iv) as the likely range of a sample, (v) as association of several variables, and (vi) as distribution (pp. 984–985).

Hence by 2007, the first 7 of the 18 chapters of Lovett and Shah's book, *Thinking with Data*, were related to variation, indicating its foundational importance before considering statistical reasoning and data analysis (Chapters 8 to 14) and learning from and making decisions with data (Chapters 15–18). The diversity of the first six chapters is typified by the page header of Chapter 7: "Variation in Variation"! Of particular interest to this study is an earlier chapter (Lehrer & Schauble, 2007), which considered two activities with children where variation first occurred in the actual measurements created by the children (e.g. estimating the height of a flagpole) and then later occurred naturally within the context of the investigation of plant growth. This research of Lehrer and Schauble, related to distribution and variation, detailed several issues encountered by Year 5 students when first introduced to the study of plant growth data. They concluded that considering variation in specific measurement data was easier for the students to deal with than considering the natural variation that occurred as the plant growth context became more complex over time. Earlier, Lehrer and Schauble (2004) had used a design methodology focussed on students' development of thinking and the support given to them in the classroom, to study their growing appreciation of natural variation through a series of experiences, including creating hand-drawn representations of data from samples of plant height. Although progressing to considering divisions in the data equivalent to box plots, the students did not have the advantage of the technology available today for representing further the variation in their data.

SOLO taxonomy: describing development of understanding

The Structure of Observed Learning Outcomes (SOLO) model (Biggs & Collis, 1982, 1991) is an hierarchical model that describes the increasing complexity of student understanding of ideas and concepts. The SOLO model has been used and extended over the years in various ways, related to the two major components of the structure: the modes of development and the advancement within each mode. First, five modes of development are proposed from birth to adulthood: the Sensori-motor mode from birth to about 18 months is described as tacit knowledge; from about 18 months the Ikonic mode emerges based on intuitive knowledge often related to visual imagery; from around 6 years, the Concrete Symbolic mode involves a transition to written language and symbol systems producing declarative knowledge; from around 14 years, the Formal mode emerges, appreciating the theoretical aspects underpinning particular disciplinary knowledge; and finally, potentially from around 20 years, the Postformal mode may develop to allow for theory creation. Of particular interest in the primary school is the transition from the Ikonic (IK) mode to the Concrete Symbolic (CS) mode.

Second, the detailed study of this development is based on the levels of advancement that take place within each mode. The first level of understanding for a mode, labelled unistructural (U), involves using a single aspect or element of the context of the mode to make a response. The multistructural level (M) involves using two

or more elements in a serial fashion in responding to a task, whereas at the relational level (R) the elements are integrated into a coherent explanation. Moving to the next higher mode involves generalising the structure into single elements of that mode. Because the CS mode is the dominant mode in primary and middle school, often observations in the previous IK mode are labelled as Prestructural and not analysed further. Groth et al. (2021), however, recently analysed IK responses to tasks, resulting in U-M-R sequences of intuitive reasoning preliminary to CS sequences. In doing so for particular tasks, some of the IK reasoning is incomplete but relevant to the context of the task and termed "normative compatible" (c), whereas other IK reasoning is irrelevant, hence termed "normative incompatible" (ic). Types of normative incompatible reasoning include a focus on irrelevant task aspects, myths and imaginative stories, out-of-context mental imaging, and a deterministic worldview (Groth et al., 2021). Again, it is possible for either normative compatible or normative incompatible responses to display U, M, or R structures, depending on how many aspects are included and how they are combined. Hence, the length of a response is not the determining factor of whether a response is IK or CS, but whether elements of the context and the specific question are employed. An IK response can be quite long but out of context, whereas a CS response, particularly at the U level, can be very short but relevant to the context. Considering responses in the IK mode in more detail, rather than just labelling them "Prestructural" as done previously (e.g. Biggs & Collis, 1982; Watson & Moritz, 1998, 2000), allows more opportunity for discussion with students on their current conceptions, why they are potentially inadequate for the task at hand, and how they might be improved. Figure 1 summarises the possible pathways for responses to tasks, noting that the IK mode is the precursor to the CS mode.

The SOLO model has been used extensively in statistics education research with school students over the years. In particular, it has been applied to learning about variation in chance contexts (e.g. Watson & Kelly, 2004; Watson & Moritz, 1998), sampling contexts (e.g. Watson & Moritz, 2000), and the study of distributions (e.g. Watson, 2009), as well as more generally across contexts based on surveys (e.g. Watson et al., 2003). More recently, the extension of Groth et al. (2021) has been applied to Year 3 students' understanding of data (Watson & Fitzallen, 2021) and to Year 5 students' appreciation of variation when working in a STEM context collecting data from catapults (Watson et al., 2022b).

Background

A 4-year project in a primary school, Years 3 to 6, sought to embed the practice of statistics and its underpinning big ideas, particularly variation, within a range of STEM-related activities. The aim was to reinforce the need to be able to use data arising across these disciplines, and particularly Science, in meaningful ways for decision-making. Each year, investigations embedded in the *Australian Curriculum: Science* (ACARA, 2019), with links to Digital and Design Technologies and other topics in Mathematics, were introduced to build understanding. Each activity from the beginning of the larger project was intended to build

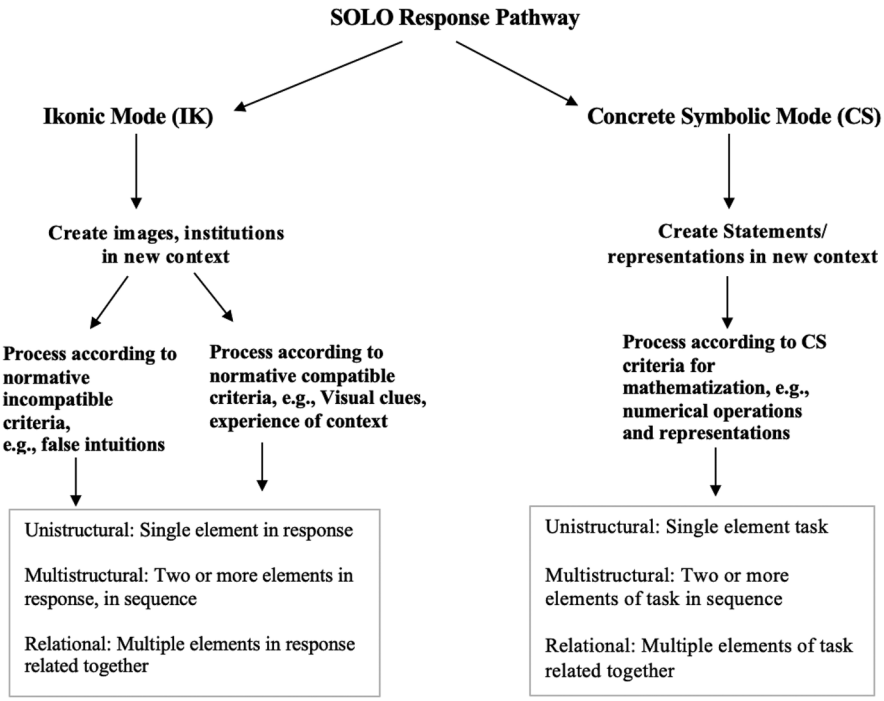


Fig. 1 SOLO model for end-of-year questionnaire items

sequentially on more complex aspects of the STEM contexts within which variation occurred and the practice of statistics was taking place. The building of the associated science concepts was linked specifically to the *Australian Curriculum* for that year of schooling (ACARA, 2019). In relation to the practice of statistics, for example, the encounter with the fundamental concept of variation (e.g. Cobb & Moore, 1997) grew first from considering variation in a measurement context (cf. Lehrer & Schauble, 2007). Initially, students experienced two different treatments with the same expected outcome (objective) but which were carried out in a different fashion: they created licorice sticks from Playdoh® by two methods, "by hand" and "by machine", but with the same measurement criteria (Watson et al., 2020b).

Variation was later experienced in a context of comparing two conditions where the expectation of outcome was also different, this time by launching ping pong balls with catapults and then changing the catapults, intending to increase the distance travelled (Watson et al., 2022b). In these two situations, class data were combined to make decisions about the research questions that were set for the students. Time was introduced as a variable in an activity involving heat, insulation, and cooling (Fitzallen et al., 2017), where again class data were combined. Each activity added to the complexity of the data analysis and decision-making because of the need to appreciate the combining of samples, the variation in the larger data sets, and the need then to find the relevant data to answer the questions.

The previous activity to the one reported here, at the end of Year 5, considered the dispersal of seeds where students worked in groups to design methods of dispersal by wind, using the engineering design process (Fitzallen et al., 2019; Smith et al., 2019). Each group of three students designed and tested one of the three methods of dispersal: helicopter, parachute, or sail. Hence in this case, all class data were not combined but instead contrasted across three subsets to answer the research question. The process involved each student in the group creating a mechanism, which was tested by the group. Subsequently, the group chose its "best" mechanism and worked to improve it further, thereby addressing the engineering design process associated with the Design Technologies curriculum. Across the class, group data for the three methods of dispersal were combined for a class discussion of the most effective method. Although the students each had the opportunity to create their original designs, the research question was based on comparing the three methods and related to the "best" method, judged by the distance travelled. The Plant Growth activity took the type of data set considered for decision-making one step further.

TinkerPlots (Konold & Miller, 2015) was a contribution from the Digital Technologies curriculum because of its particular features assisting students to display and understand variation in distributions (Konold, 2007; Konold & Lehrer, 2008; Konold et al., 2007). Watson and Fitzallen (2016) examined the software from the perspective of affordances (Chick, 2007; Gibson, 1977), in that it had the capability to foster learning based on its accessible and flexible features, facilitating its incorporation in learning activities by teachers. They provided examples of affordances related to students' learning of statistics, to enhancing teachers' pedagogical content knowledge for teaching statistics, to assessing students' understanding, and to supporting other areas of the curriculum. In the case of this study, that other area of the curriculum is obviously Science. Successful implementation of learning activities supported by *TinkerPlots* across the school years has been reported in recent years (e.g. Allmond & Makar, 2014; Fielding-Wells & Hillman, 2018; Kazak et al., 2014; Khairiree & Kurusatian, 2009).

Before the Plant Growth activity, students had learned to create plots in *TinkerPlots* when data were provided with the Licorice activity, to enter their own data with the Catapults activity, and to display numerical and multiple categorical variables with Wind Dispersal of Seeds. Through an activity on Viscosity (Watson et al., 2022a), they also plotted the relationship of two numerical variables. A summary of these activities is found in Fitzallen and Watson (2020). This report presents some outcomes from the final major activity in the project, Plant Growth.

The Plant Growth activity then focused on combining the statistical understanding gained with the Science Inquiry Skills to allow students to decide, within the context of plant growth, their own research question, collect the appropriate data, represent and analyse the data in *TinkerPlots*, and present their conclusions with respect to the method of plant growth chosen.

Building upon interest in plants, the Year 6 *Australian Curriculum: Science* (ACARA, 2019) suggested the following Content Description and Elaboration for Biological Sciences:

The growth and survival of living things are affected by physical conditions of their environment (ACSSU094)

- Investigating how changing the physical conditions for plants impacts on their growth and survival such as saltwater, use of fertilizers, and soil types.

Plant growth has been suggested often as a learning context in the classroom. As early as 1984, Smith and Anderson explored the teaching strategies employed to support students' conceptual understanding of plant growth. Wood and Roper (2000) suggested a basic technique for students to measure the number of expanded leaves for garden bean plants grown in the classroom under three lighting conditions over an 11-day period. The *GAISE Reports* (Bargagliotti et al., 2020; Franklin et al., 2007) provided a similar example by considering growth data from radish seedlings grown in three conditions of darkness and light. In line with the *Australian Curriculum* (ACARA, 2019), the Australian Academy of Science (2016) produced a 7-lesson resource for Year 6 focusing on sustainability, water, salt, and plant growth. One lesson considered the growth of lettuce seedlings in four different saltwater concentrations. An optional lesson included students working in teams to formulate a question related to salinity and conducting an investigation to collect evidence to support or refute their claims. These reports did not include any research outcomes related to student learning as a consequence of carrying out the experiments.

Watching plants grow involves observing variation: they become taller, produce more leaves, flower, bear fruit, and/or produce seeds with different dispersion mechanisms. Plants change across the seasons and eventually they may shrink and die, or become dormant until the weather is again favourable for producing flowers and fruit. The aspects of plant growth that occur on the germination of seeds provide a context in which students can observe change and variation over time. As a Biological Science, the study of plant growth can contribute to appreciation of the importance of variation not only in Science but also across the STEM disciplines. At the school level, experimenting with plant growth, as reported here, was aimed at linking content expectations across the areas of the curriculum associated with Science, Technology, Engineering, and Mathematics. Although in most countries, Science and Mathematics are distinctive content areas of the school curriculum, often Technology and Engineering are not singled out individually. In Australia, for example, these topics are considered under *Digital and Design Technologies* (ACARA, 2019), whereas in the USA, Engineering is considered by the National Research Council (NRC) as part of the Science curriculum (2013). In particular, based on the *Australian Curriculum* (ACARA, 2019), this study focused on Year 6 Science Understanding (Biological Sciences) and Science Inquiry Skills from Science; Measurement and Statistics from Mathematics; data handling software from Digital Technologies; and experimental design and testing from Design and Technologies.

The overall research interest and questions in this study related to the features associated with variation that students included when using *TinkerPlots* software to interpret plots of data collected from a plant growth experiment. Initially, this related to the learning activity as it occurred, with students reporting in their workbooks the

analyses they were carrying out to explain the variation they were seeing. Later in an end-of-year questionnaire, the focus was on students' consolidation of understanding of variation in terms of the SOLO developmental model as they answered questions about the activity.

Methodology

The methodology for the Plant Growth study followed a pragmatic paradigm (Mackenzie & Knipe, 2006) in being problem-centred and oriented to real-world practice with specific questions matched to the purpose of the research. Here, the initial data were qualitative in nature, with the aim of capturing the nuances of students' *formative* thinking (e.g. Wiliam, 2011) while they were carrying out the Plant Growth activity. Second, again following Wiliam (2011), data were collected to monitor the *summative* learning as a consequence of the activity. The method hence was based on a classroom intervention—the problem-centred, real-world experimental activity, with accompanying workbook questions—followed by an extended end-of-year multipart question designed to monitor learning.

In the classroom: the experiment

The Plant Growth activity was a variation on that suggested by the Australian Academy of Science (2016). It was decided that within the context of an appropriate experimental design, groups of four students would themselves decide the research question and hence the materials to be used in setting up the experiment with a control and three treatments using an hydroponic watering system (Fig. 2). There were four soil types, four treatments, and two different seeds. Each group could choose a seed, a soil type, and a treatment, or test the influence of the various soil types (see Appendix 2 for details). Hence, it was very unlikely that in a class there would be a duplicate design. Each group of students was its own scientific team, carrying out

Fig. 2 Planting wheat seeds in the hydroponic bottle system and growth after 14 days

the instructions for setting up the experiment, with each student supervising one of the treatments. Measurements were taken across 3 or 4 weeks, including recording the data, entering the data into *TinkerPlots*, and making a decision about the plants' growth in relation to the treatments chosen to answer the question: What makes plants grow best?

The purpose of allowing a large degree of freedom in the choice of materials and treatments was to introduce students to the wide range of possibilities and to the importance of data when applying Science Inquiry Skills. As the activity was set up, it was not expected to be possible, at the end, for the students to all agree on "the best" of various treatments across the class; some treatments, however, were agreed not to be appropriate.

The details of setting up and carrying out the experiments are presented in Wright et al. (2021). This report considers two aspects of the extended activity. When the experiments were complete, students were given instructions for entering their data into *TinkerPlots* and each member of the group had access to the complete data collected by the group. An example is shown in Fig. 3, which involved planting wheat seeds in potting mix with the treatment being three different concentrations of vinegar in the water. In their workbooks, students then answered the questions in Fig. 4 (condensed, with space for writing removed), which focused specifically on the variation experienced during the activity as seen in the *TinkerPlots* graphs, within and across the variables. The instructions were very explicit due to the complexity of the experiment and the students' lack of previous experience. The data from these workbook questions provide the data for the formative assessment.

Longer-term learning outcomes

At the end of the year, students completed an extended questionnaire on the year's activities and other linking items across the 4 years of the project. The first multipart question was related to the Plant Growth activity and is shown in Fig. 5. As can be seen, one example of a *TinkerPlots* graph was presented and students were

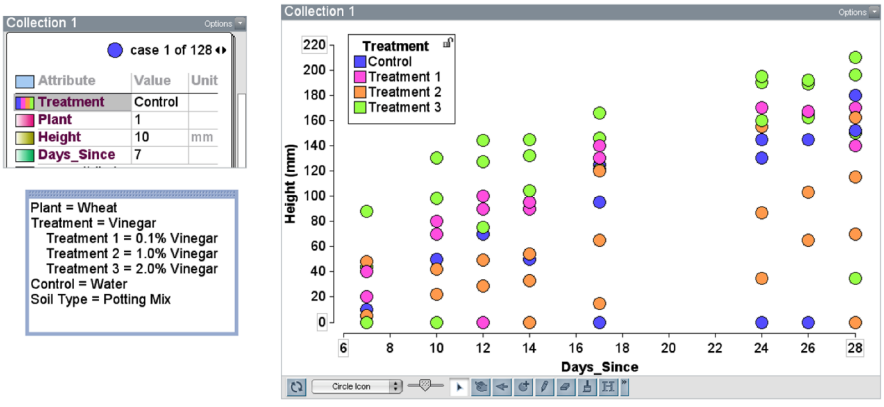


Fig. 3 *TinkerPlots* file for one group's data

Analyse the Data

Use the plot (graph) you created to answer these questions.

1. **Identify the variation**
   Describe 3 different ways variation is shown in your plot.

2. **Analyse the data for one of the Treatments**
   Choose a treatment (tick one)
   - ☐ Control
   - ☐ Treatment 1
   - ☐ Treatment 2
   - ☐ Treatment 3

   What does the variation *within* that treatment tell you?

3. **Analyse the data comparing two of the Treatments**
   Choose two treatments (tick two)
   - ☐ Control
   - ☐ Treatment 1
   - ☐ Treatment 2
   - ☐ Treatment 3

   What does the variation *between* those two different treatments tell you?

4. **Analyse the data according to the days grown**
   What does the variation across the **Days_Since** attribute tell you?

5. **Analyse the data in the plot**
   What conclusions do you draw about the *typical growth* of the plants?

Fig. 4 Workbook questions focussing on variation (space for writing removed)

asked to explain and interpret the outcomes from it. This was to evaluate the summative aspect of learning from the activity. Because of the time gap from completing the activity for some of the students, the initial questions were intended to be straightforward, helping students recall the main components of the activity before discussing the outcomes.

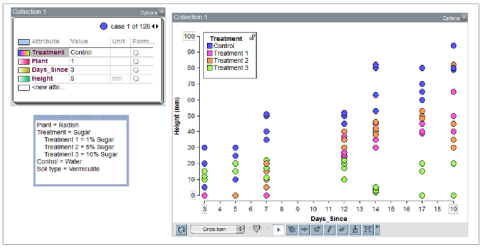
Research questions

In order to appreciate the two aspects of learning that took place in relation to the Plant Growth activity, it is necessary to document the students' formative experiences while taking part in the experiment, as well as the summative results from the end of the year (Wiliam, 2011).

Fig. 5 End-of-year multipart question (space for writing removed)

STEM: Modelling with Data Student Questionnaire 2018

Q1. Here is a plot of the data from one group's experiment on plant growth.



(a) Why did we have a control and 3 treatments?
(b) What data did we collect to plot?
(c) Why is there a gap in the data?
(d) With your pen or pencil, draw a line through the data for the control. Now write a sentence on how the data changed.
(e) How did the data for the control compare with the other treatments?
(f) Overall, what did this group learn from their experiment?

Research Question 1 (formative) *What were the explicit features associated with variation that students included in their descriptions of a science experiment on plant growth (cf.* Fig. 4*) based on their groups' TinkerPlots representations?*
Research Question 2 (summative) *What evidence of student learning associated with variation in the context of the Plant Growth activity was demonstrated in responses to the end-of-year question based on one group's data (cf.* Fig. 5*)?*

The end-of-year multipart question presented all students with a single example of the outcomes of an experiment in the form of a *TinkerPlots* graph (cf. Fig. 3). Although students had seen their classmates give oral presentations on each group's working, they only had first-hand experience within their own groups. The second research question hence relates to the consolidation of learning and application to a particular experimental data set. The word "variation" was purposely not included in the end-of-year question.

Participants

Three classes of Year 6 students at an urban independent Catholic school participated in the study with 64 of the 71 students having parental and student permission for their data to be included in the research reported here. The school, in an inner regional center, had a socio-economic status index (ICSEA) value of 1026 (mySchool.com.au; mean = 1000, standard deviation = 100) and four of the students had English as a second language. The reported sample consisted of 40 boys and 24 girls, ranging from 11 to 12 years of age. Each class was taught by the same teacher, using teaching notes and student workbooks developed by the research team, with input from the teacher. The activity took place across school terms 1, 2, and 4. Eight of the 64 students were unavailable for the end-of-year questionnaire, which was administered at the end of term 4 and the entire study. The overall longitudinal study

for which the Plant Growth activity was the final part had ethics approval from the Tasmanian Social Sciences Human Research Ethics Committee (approval number H0015039).

Analysis for Research Question 1

As the first research question was formative in nature (Wiliam, 2011), the data analysis was considered qualitative and interpretative (Creswell, 2013). As noted, the experiment completed by each group of students was unique with its own set of data to interpret. Furthermore, the activity took place across three school terms for three Year 6 cohorts, often with different climatic conditions under which the plants were growing. It was hence important to assess the types of response without judgement on the variable/s chosen or the experimental conclusions drawn for growth. All responses to the questions in Fig. 4 were clustered into categories by the first author related to the expectations as stated specifically in each of the workbook questions. The same process was carried out by the second author independently, with a discussion confirming the observed categories, which are summarised in Table 1.

Analysis for Research Question 2

For the second research question, because the end-of-year question was monitoring the summative learning (Wiliam, 2011) that had taken place in relation to the Plant Growth activity, the Structure of Observed Learning Outcomes (SOLO) Model (Biggs & Collis, 1982, 1991), as adapted for the Ikonic Mode (Groth et al., 2021), was appropriate for categorising the outcomes.

Table 1 Categories of response to workbook questions

Workbook question	Observed responses
(1) Describe 3 different ways the variation is shown in your plot.	• List of the 3 variables • Description of: - Variation of 1 variable - Variation of 2: height & treatment - Variation of 2: height & days • Extended description of all 3 variables
(2) Choose a treatment. What does the variation *within* that treatment tell you?	• Statement "outside" of treatment chosen • One, two, or three comments about the treatment chosen
(3) Choose 2 treatments. What does the variation *between* those two different treatments tell you?	• Response not related to the data • One, two, or three comments on the two treatments chosen
(4) What does the variation across the **Days_Since** attribute tell you?	• One statement • ≥ Two statements • One comparison • ≥ Two comparisons
(5) What conclusions do you draw about the typical growth of the plants?	• Comment/s not related to the experiment • Comments on treatments • Generalisations to context

In considering the coding of responses to the end-of-year questionnaire items, the nature of the individual questions targeted different levels of response in the CS mode. In particular, items (b) and (c) were included to reorient the students to the activity. This was taken into account when rubrics were developed for coding responses. There were also many potential elements of the task to be used in responses due to the complex nature of the overall activity. Some may or may not have been relevant to the particular question being answered. These elements included the physical components of the experiment (e.g. soil type, treatment kind, plant type); the actions taking place (e.g. watering, growth, wilting, measuring, comparing); and the products of the investigation (e.g. data, plot, conclusion).

Table 2 details the notation used to describe the modes and levels of responses employed here for responses to the Plant Growth multipart question in the end-of-year questionnaire. CS responses are those where elements relevant to the question are included. For these responses, the difference between M and R levels is often the use of a word such as "because", which shows the linking between or among the elements listed. IK responses may be vague or intuitive with unclear or unspecific relation to the question asked. If responses are considered within the context of the particular question, they are deemed normative compatible (c). If they are considered irrelevant, or incorrect with respect to the context of the question, they are deemed normative incompatible (ic). As an example, for the question about why there was a gap in the data, a normative incompatible response would give a reason not related to the gap ("to measure") or an irrelevant reason ("because different plants grow different"), whereas a normative compatible response might be related to the gap ("because of the numbers on the line"). Table 2 is used as a rubric for coding the end-of-year questionnaire.

Coding was undertaken by the first two authors. The overall agreement rate for the student workbook items was 85%, ranging from 66% for Q2 to 98% for Q5. For coding of the student summative assessment items, the agreement rate was 96%, with a range from 93 to 100%. All disagreements were resolved by discussion.

Results

The results are presented in two parts. First, the categories identified in the responses to the five workbook questions (Fig. 4) are identified with examples for answering *Research Question* 1, related to the formative aspect of learning. Second, the SOLO analysis of the responses to the end-of-year questions (Fig. 5) is presented with examples for answering *Research Question* 2, related to the summative aspect of the learning.

Table 2 SOLO modes and levels of response

Mode		Description	Level	Description
IK	IK(ic)	Focus on irrelevant, out-of-context, or incorrect aspect of the question	U	Single comment $IK(c)_U$ or $IK(ic)_U$
	IK(c)	Intuitive, idiosyncratic, or nonspecific reference in the context of the question	M	Two or more aspects in sequence $IK(c)_M$ or $IK(ic)_M$
			R	Two or more aspects related together $IK(c)_R$ or $IK(ic)_R$
CS		Declarative response (language, symbols, and representations) using elements provided in the task to respond to a specific question	U	Single element employed in response to the question CS_U
			M	Multiple elements employed in sequence to respond to the question CS_M
			R	Multiple elements integrated in response to the question CS_R

Workbook responses for Research Question 1: what were the explicit features associated with variation that students included in their descriptions of a science experiment on plant growth (cf. Fig. 4) based on their groups' *TinkerPlots* representations?

Q1. Describe 3 different ways variation is shown in the plot

This question could be answered succinctly by listing the three different variables labelled in the plot: Height, Treatment, and Days_Since (referring to the days since planting). This was done by 20% of students, who elaborated no further. Other students (23%) went further in the description of the "ways" these three variables represented the variation seen in the plots. At times this was an extended description of one of the variables, such as "different amounts of salt" [ID175] or "plants went up then most went back down and then they went up again" [ID117]. Nine students (14%) commented on a single variable only, with eight commenting on Height and one on Treatment. Three students (5%) did not respond to this question.

Most frequently, the variation was described between the two variables, Height and Treatment, often with implicit language for height (e.g. tallest) and comparison (31%), as shown in the following examples.

- Treatment 1 grew the tallest. [ID166]
- The heights are different. Control got in the hundreds and treatment 2 didn't grow at all. [ID138]
- With the differences of control and treatment 1, 2 & 3 control has grown the most but in some places they are very consistent. [ID119]
- There is variation the plants grew for the different treatments. [ID135]
- Most of the control is on 1 mm and Treatment 1 has the highest TinkerPlot and treatment 2 and 3 are the same really. [ID159]

Although time (Days_Since) displayed in the plot influenced many of the comments made about *TinkerPlots*, it was specifically mentioned less frequently in conjunction with height (6%). The examples that follow illustrate the types of connections made.

- Most of the variation was in between 19 days, with the height there is also variation between the days 12 and 14. [ID148]
- I think the plants have grown a lot over the time when we first planted them. Mine has grown 130 mm … [ID170]
- [It] goes from 0 to 220 and there is 20 days. [ID115]

Q2. Analyse the data for one Treatment. Choose one. What does the variation within the treatment tell you?

The appropriate responses typically made were one or two general comments about the plants' growth in relation to height for the treatment chosen, although eight students (13%) linked three comments together. Examples of these are shown in Table 3. Because the question related to a single treatment, the comparison of

Table 3 Variation within treatment*

One comment (19%)	Two comments (39%)	Three comments (13%)
The data says that the plant is *getting taller*. [ID103] Everytime that we measured, our Control *grew bigger*. [ID123]	It gradually *grows* then *shrinks*. [ID121] At first they *didn't grow* but then they *shot up*. [ID162]	That the first measurement was *below 20 mm*. It gradually *moved up* and the highest was *only 160 mm*. [ID109] My 4 plants are *different* level to each other. One will be *higher and* the others will be *lower*. Past few days my plants grow *higher and higher*. [ID156]

*Five students (8%) did not respond

the treatment with another treatment was not appropriate. This accounted for 22% of the comments, and included:

- That Treatment 1 is bigger than all of the others. [ID143]
- Control had no sugar and others did and control always was the tallest. [ID127]

Some students combined a within-treatment statement with a comparison statement. These responses were given credit for the appropriate comment.

Q3. Analyse the data comparing two treatments. What does the variation between those two treatments tell you?

Generally, responses pointed out one or two differences or similarities between the two treatments (69%). Very few took the comparison outside the two treatments chosen. Occasionally, comments (13%) went outside of the data available in the plots when suggesting idiosyncratic causes for difference. These included:

- Maybe treatment 3 had more sun and a good amount of water and maybe treatment 2 had too much water and not enough sun. [ID130]
- That our prediction was incorrect. We believed that the vinegar being acidic would kill the plant because a young plant's roots wouldn't be able to handle the vinegar's acidity. [ID136]

Table 4 contains examples of some of the most common responses.

Q4. Analyse the data according to the days grown. What does the variation across the Days_Since attribute tell you?

As time was the essence of the experiment on plant growth, it provided the greatest opportunity to discuss changes in the other variables. With the focus placed on the plot, responses generally recounted observations drawn from there in terms of the variables Treatment and Height, rather than explanations on the science of the plant growth (see Q5). Only one response did not make a connection to the other variables: "Most of the *TinkerPlots* are on 10–14" [ID159], and two others were unrelated or idiosyncratic (5% in total). Following on from Q3, some responses continued to refer back to the previous question. For example, a response of "Control was the highest" for comparing Control and Treatment 2 was followed by "That they went higher." [ID115] when referring to Days_Since. These responses were considered appropriate. The responses ranged from one or more generalised statements, to one or more contrasts observed for the other variables, as exemplified in Table 5.

Table 4 Variation between treatments*

One comment (31%)	Two comments (38%)	Three comments (3%)
The treatment 3 had more sugar and the *plant didn't like it*. So it didn't grow much. [ID126]	*At first treatment 2 was higher* for like the first 3 times *then control shot up* and growed [sic] more higher. [ID173]	That they were different but both *grow well*. Some were above 200 but some *stayed* in 180's. Treatment 1 grew a *bit less* than control. [ID152]
The worm casting does well but *compost is better* because it has minerals and vitamins. [ID164]	Treatment one *started bigger* than control and has much *bigger changes*. [ID168]	The control is *very tall* but treatment 2 is the *second highes*t so in some places they are *the same* but in others they are *very different*. [ID119]
The data says that the plant is *getting taller*. [ID103]	Control is *the tallest*, it is higher than 200 but Treatment 3 is *the second tallest*. But it is inbetween the 100. [ID131]	
[Treatments 1 and 2] They are *pretty similar* in height. [ID103]		

*10 students (16%) did not respond

Table 5 Variation across Days_Since*

One statement (16%)	≥ Two statements (5%)	One comparison (31%)	≥ Two comparisons (28%)
It tells me that plants *grow* as time passes. [ID121] How far the plants *grew* throughout the days. [ID102] They have *changed* a lot since day 7. [ID162] That some plants have been *growing* lots over the days. [ID165] That they are *growing* taller and taller. [ID146] How far the plants *grew* throughout the days. [ID102]	They got *higher* through each measurement and each week *from* 20 *to* 160 mm and 60 mm to 200 mm. [ID109] There was quite a bit of *difference* with height. It *grew* a few millimetres each day. [ID150] That even at the last day we me[a]sured that the plants have *grown* a lot. That the treatments with sugar the longer the sugar was in the water *didn't help* the plant when growing. [ID149]	Some have grown *more and* some have grown *less*. [ID126] So T3 was in the *lead* at the start but at the end it came *second highest*. [ID104] The control gets *bigger and bigger* but the changes are *less rapid* than the others. [ID168] Everything grew *pretty similar* except treatment 2. [ID127] I think some of the plants have *grown fast* but some have *not grown that fast*. [ID131]	That some plants [grew] *higher and* some [grew] *shorter*. Over time they got higher but it *took longer* for others. [ID116] It tells me that the *size varies* through the days. The plant growth was *unsteady* and treatment 3 sprouted up *very fast* and 2 sprouted very *slowly*. [ID174] On day 7 they're all *very distant* but as the days go up the plants slowly start to *come closer* but on day 19 and 21 start to *go distant again*. [ID114]

*10 students (16%) did not respond

Q5. Analyse the data in the plot. What conclusions do you draw about typical growth of the plants?

Because there were so many possible designs (17) for the experiments with different expectations, the responses to the final workbook question ranged widely over the possibilities. Some (33%) stayed focused on the data in the plots, revising the outcomes for various treatments.

- Treatment 3: It grew slow but well. Treatment 2: Did not grow too well but it was OK I guess. Treatment 1: It grew well and strong. Control: It grew the best out of all of them with just water. [ID117]
- That through all the measurement, the plants got higher and higher each week. Different things, like vinegar, etc., can make a big big difference. [ID109]
- Treatment 2 and 3 grew the best. Treatment 1 and Control didn't grow as well. [ID165]

Other responses made generalisations from their group's experience (20%).

- Sugar is not good for plants. It can make them grow but only for a matter of time, before it kills them. My plants I say grew the best because [they] had no sugar. [ID105]
- If you have too much sugar it grows a lot and if you just have water it also grows a lot but if you have too less or not enough sugar your plant may not grow as well as the other plants or not even grow at all. [ID114]
- It doesn't need anything added, it just grows naturally because control which had no sugar grew the best. [ID127]

Some responses, however, reflected less appreciation of the finer points of the experiment for typical plant growth (22%).

- You put water in it and it will go up so that's how to grow a plant. [ID115]
- I have learnt that plants die out. [ID147]
- I think that all plants slowly grow. [ID121]
- That even if it was the same thing to grow, the data is always different and never the same and it just depends on luck if it grows well or not. [ID116]

One student made a comment peripheral to plant growth and 15 (23%) did not respond to the question.

End-of-year responses for Research Question 2: what evidence of student learning associated with variation in the context of the Plant Growth activity was demonstrated in responses to the end-of-year multipart question based on one group's data (cf. Fig. 5)?

Using the rubric described in Table 2, each summative assessment question (see Fig. 5) is considered in turn. It should be noted that the construction of the questions at the end of the year was designed to assist the students in remembering the parts of the activity and how they were linked to the plot provided rather than necessarily expecting a full range of SOLO responses.

(a) Why did we have a control and 3 treatments?

Generally, students showed a good appreciation of why a control and three treatments were required for their experiments with only one student providing a normative incompatible response and 85% of students focusing on the essential elements of variation, i.e. comparison and difference, for CS responses (see Table 6). Half, however, only noted one specific reason.

(b) What data did we collect to plot?

Appreciation of variation in responses to this question was reflected in the number of different variables measured, as well as the implicit appreciation of the change associated with growth. Again, only one response was considered normative incompatible (see Table 7). Given the specific nature of the focus on the measurement of "data" in the activity, however, only half of the students were specific about what was measured or noted a single variable or list of the variables, e.g. Height, Treatment, Days_Since. Forty-seven percent of responses were ikonic in referring more obliquely to "measurement" or "growth". The split of responses such as this illustrates the value of acknowledging IK responses as less focused on the actual "data" than CS responses but still appreciating the context in which the question is asked.

(c) Why is there a gap in the data?

This question was included to check if the awareness of missing data (see Fig. 5), which had been discussed pointedly with the classes, had been internalised (see Table 8). As it turned out, each class had a reason for absence during the activity, being the Easter break, a trip to the nation's capital, or a school camp. At the time of entering data into *TinkerPlots*, the importance of acknowledging the data were not collected at particular times was emphasised and hence there were gaps with no icons for some values on the *x*-axis. Eighty percent of students recalled this aspect of the data entry, whereas 18% of responses were considered normative incompatible because they did not recall the context of the experiments.

Table 6 SOLO levels for "Why did we have a control and 3 treatments?"

Mode	Level		*n*	%
IK	$IK(ic)_U$	We required control to control the 3 treatments. [ID147]	1	2
	$IK(c)_U$	To evaluate all the options. [ID171]	4	7
	$IK(c)_M$	To put the data in the plots to show the results. [ID130] Because there was 3 other people in my group and everyone has a different amount of an ingredient. [ID104]	2	4
	$IK(c)_R$	Control has no sugar which is the treatment which means the other 3 treatments have different amounts of sugar. [ID161]	1	2
CS	CS_U	To compare the data. [ID117] To see different outcomes from each treatment. [ID135] To see which one would work better. [ID139]	30	54
	CS_M	I think the reason why is so you can see the variation or see if the experiment has worked. [ID105]	6	11
	CS_R	We had 3 treatments because we wanted to see which one would grow the furthest in different types of soil. [ID141] So we could see the variety and whose plant grew differently with the different treatments and control. [ID158]	11	20
No response			1	2

Table 7 SOLO levels for "What data did we collect to plot?"

Mode	Level		*n*	%
IK	$IK(ic)_U$	Our plant that we made. [ID146]	1	2
	$IK(c)_U$	We measured them. [ID108] Plant growth. [ID157]	15	27
	$IK(c)_M$	We collected plant growth and put it in TinkerPlots. [ID160] How plants grow and what amount of salt grows best. [ID124]	11	20
CS	CS_U	How high they were! [ID129] How tall they grew. [ID135]	15	27
	CS_M	Treatment, plant, days since and height. [ID101] How high the plants would grow and also how many days it took. [ID155]	13	23
No response			1	2

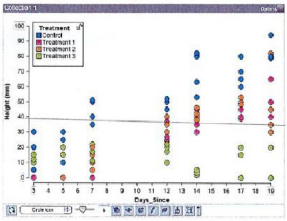
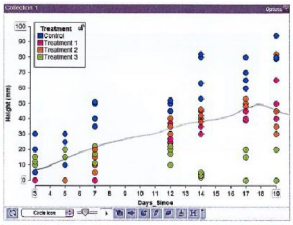
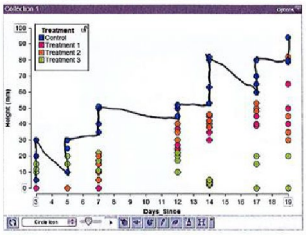
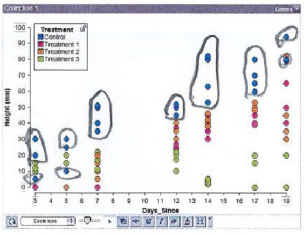
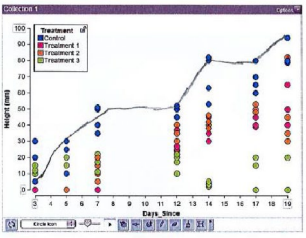
(d) Draw a line through the data for the control

This was a task about summarizing visually the variation in the data over time. To complete the task, students needed to read the key in the plot to determine which data were the control data (see Fig. 5). This was the basic criterion for a CS response. Ninety-two percent of students managed this successfully and their "lines" drawings were considered to be in the CS mode, representing the upward trend in the data (see Table 9). Four others did not recognise the control. Of these, two were considered IK incompatible, as showing no relationship at all, whereas two others were considered IK compatible by showing a trend but not related to the control. Although many of the CS responses were not considered "lines" as would be accepted in more senior classrooms, they identified the

Table 8 SOLO levels for "Why is there a gap in the data?"

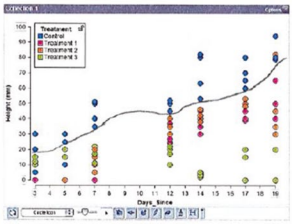
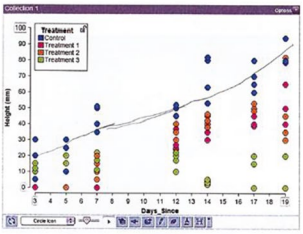
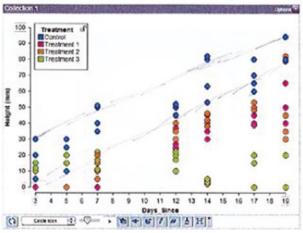
Mode	Level		*n*	%
IK	$IK(ic)_U$	To measure. [ID102] To show each treatment. [ID115] Because we only had to measure it on the days our plant improved. [ID141]	8	14
	$IK(ic)_M$	Because different plants grow different to the rest and some grow higher than others. [ID154] To know how tall is that treatment and what day it is. [ID159]	2	4
CS	CS_U	We did not record on those days. [ID121] The two days each week we did not measure. [ID123] We went to Canberra. [ID131]	40	71
	CS_M	Maybe we were away or forgot to measure the plants. [ID101] Because we didn't measure that days. We were away. [ID167]	5	9
No response			1	2

Table 9 SOLO levels for "Draw a line through the data for the control."

Mode	Level	Description	Plot	*n*	%
IK	$IK(ic)_U$	Line no relationship to the data. [ID106]		2	4
	$IK(c)_U$	Line not related to the control data; idiosyncratic drawing. [ID148]		2	4
CS	CS_U	Continuous line connecting all control data points or otherwise recognising control data; e.g., broken lines or "circles" around data. [ID161] [ID147]	 	8	14
	CS_M	Continuous bumpy line through maximum and/or minimum of control data. [ID103] [ID102]		26	46

required data, the control, which was the context for the question. The unistructural ones highlighted in some fashion all of the control data, whereas the multistructural responses provided a continuous line connecting "boundary" values of the control data, and the relational responses attempted to show a general trend with a "smooth" line within the control data or suggesting smooth boundaries.

Table 9 (continued)

Mode	Level	Description	Plot	*n*	%
					
	CS_R	Continuous indication of smooth trend upward the middle of control data, or similar "range" lines around control data. [ID114] [ID104]		16	29
					
No response				2	4

(d$^+$) Write a sentence on how the data changed

For this question on *how* the data changed, 81% of the students were able to describe with CS evidence the variation in the data, with a few integrating their explanations at the relational level (see Table 10). Again, there were a few responses (16%) that were either compatible or incompatible with the context of change and not focused on the "how" aspect of the question.

(e) How did the data for the control compare with the other treatments?

The focus of this question was a comparison of the control with the other treatments, with 88% of students able to make at least one reasonable description, using language describing variation, e.g. consistency, taller, different, bigger, faster. Of the five responses (10%) that did not focus on the data (see Table 11), four were, however, compatible with the context of the question.

Table 10 SOLO levels for "Write a sentence on how the data changed"

Mode	Level		*n*	%
IK	$IK(ic)_U$	Only 3 treatments. [ID115] It didn't really change. [ID143] Different groups of control. [ID147]	4	7
	$IK(c)_U$	By each day something different will happen every day. [ID146]	2	4
	$IK(c)_M$	The data changed because we had the plant by the sun and it was getting watered every second [day] and we had different soils. [ID173]	3	5
CS	CS_U	It went higher. [ID150] It grew over the period of time. [ID165] It grew taller over each week. [ID135]	21	38
	CS_M	It started off smooth then it went up really fast. [ID112] The control was always growing up but at the start one of them was second bottom. [ID141] It is very inconsistent but going up. There's a lot of variation in the way it goes up-straight-up-straight. [ID101]	20	36
	CS_R	The data went up and up and that shows that doing it normally is the best way. [ID111] The control grew then stays the same for a few days then shoots up again and it keeps going on like that. [ID138]	4	7
No response			2	4

Table 11 SOLO levels for "How did the data for the control compare with the other treatments?"

Mode	Level		*n*	%
IK	$IK(ic)_U$	More because it had time to grow. [ID146]	1	2
	$IK(c)_U$	Good, better. [ID134]	1	2
	$IK(c)_M$	The control only has water and no treatment but treatment 1, treatment 2, treatment 3 all do. [ID169]	2	4
	$IK(c)_R$	That depends what you had in your water that made your plant grow or die, some died because they had sugar or others. [ID154]	1	2
CS	CS_U	Different. [ID150] It grew the tallest. [ID123] Got higher because it was grown like a normal plant. [ID139]	30	54
	CS_M	Control grew more consistently and did not go up and down. [ID121] Control grew the most and like treatment 3 didn't grow a lot, it's down the bottom of the chart. [ID138] It was always at the top of the plots and it was almost colour coordinated and it's almost in order. [ID112]	13	23
	CS_R	Grew very well and got very high. It is the least consistent of all the treatments but is [probably] grew the normalest because it didn't have any thing added to the water. [ID101] Control (on this data) is growing much bigger and faster than all the other treatments. So that means the method control used is the best one to use for your plant. [ID167]	6	11
No response			2	4

(f) What did this group learn from their experiment?

This question was open-ended and some students interpreted it from their experience more generally rather than from the example given in Fig. 5. Sixty-eight percent responded in line with the expectations of learning about variation embedded in the extended activity, although not explicitly using a related word. Twenty-eight percent made IK suggestions, with 16% considered incompatible with the learning objectives (see Table 12).

Discussion

Representing, analysing, and interpreting the variation in the data from the Plant Growth activity that occurred at the end of each of the three classroom interventions (cf. Fig. 4) formed the context for Research Question 1 of this report. The work reported was as a culmination of other workbook entries and many discussions that took place across the term as the experiments were set up and monitored over the three or more weeks of gathering data. There was often debate in the classroom about the expectations and surprises related to the emerging results for the various treatments. Each group of four also presented a report to the class with questions and discussion after completing the workbook questions (see also Wright et al., 2021).

Specifically, the classroom intervention part of this study illustrated six of the conceptions of variability suggested by Shaughnessy (2007): (i) variation was noted

Table 12 SOLO levels for "What did this group learn from their experiment?"

Mode	Level		*n*	%
IK	$IK(ic)_U$	Some plants grow in small vinegar amounts. [ID115] Sugar can help grow plants. [ID104]	9	16
	$IK(c)_U$	Different stuff grow better than other thing. [ID146] That you need to measure each day. [ID150]	4	7
	$IK(c)_M$	That it grows every day and we worked as a team. [ID106] Some plants will grow very well and some won't do so well. [ID114]	3	5
CS	CS_U	How plants grow. [ID159] Data isn't always consistent. [ID161] What treatment works best. [ID109] That no sugar grows more. [ID132]	23	41
	CS_M	What plants do and don't like, what they need to grow, how to record data, make graphs, read graphs. [ID101] That sugar isn't good for radish and that you should use regular water instead. [ID149] That people's plants are different at growing and that if you use different things plants will grow differently. [ID131]	11	20
	CS_R	That plants can grow with different substances but die if given too much different liquid suited to its adaption. [ID126] That not all treatments work or complete due to the treatment type or the time period. [ID174]	4	7
No response			2	4

in particular values as students collected and recorded their data. (ii) Variation as change over time was built into the activity and perhaps was the most obvious aspect to students as they collected data from their plants on many occasions, commenting on changes in weather conditions or growth spurts. (iii) Variation as the whole range of possible values was experienced through the *TinkerPlots* representations. (iv) Variation as the range of possible samples was clear as the students within their groups discussed and compared the different treatments being applied to the plants. (v) Variation as association of several variables was seen through the plots of Days_Since and Height with different colours for the different treatments. (vi) Variation as distribution was seen as the class observed and discussed the results of other groups.

The expectations of the workbook questions based on the *TinkerPlots* representations the students created were quite advanced for Year 6 students. Although from Year 4, the *Australian Curriculum* (ACARA, 2019) Content Descriptions related to "Data representation and interpretation" suggest students "evaluate the effectiveness of different displays in illustrating data features including variability" (ACMSP097), the graphical contexts do not include more than two variables. Across Years 4 to 6, suggestions include, "construct suitable data displays … [that] include tables, column graphs and picture graphs where one picture can represent many data values" (ACMSP096); "construct displays including column graphs, dot plots and tables, appropriate for data type …" (ACMSP119); and "interpret and compare, a range of data displays, including side-by-side column graphs for two categorical variables" (ACMSP147). The expectations associated with the graphs these students created went well beyond this. The more statistical language of "within" and "between" variables is not used in the curriculum. The Proficiencies in the *Curriculum* for Statistics (Appendix 1), particularly "reasoning" and "problem-solving," cover well the general aims across Years 1 to 6. There needs to be, however, more explicit linkage of these to the details in the Content Descriptions.

The only previous classroom research that could be found in the context of collecting data associated with plant growth was that of Lehrer and Schauble (2004, 2007). As a design study, their research documented the progression from the beginning of the students' experiences with variation in measurement data and representing it in hand-drawn displays. As the students in the current study had experienced 3 years of discussion about, and representing of, variation, often in hand-drawn graphs, they were beginning the activity with a more sophisticated background and experience with the software *TinkerPlots*. It was hence felt important to begin with what the research team felt were significant questions for the students to answer during the activity. Although not ideal that some of the workbook entries were incomplete, the teacher said that the expectation of writing detailed explanations was quite demanding for Year 6. The teacher, however, was favourable to the approach and later had the students prepare other written reports for their classroom assessment. The research team was satisfied that generally speaking over half of the responses to the workbook questions were acceptable, showing appreciation of the meaning of "within" (71%), "between" (72%), and "across" (79%) in relation to the variables and the graph presented (cf. Tables 3, 4, and 5), indicating appreciation of the concept of variation and the role it played in making decisions about what makes plants grow best. Overall, the students were able to answer their research questions

and were able to recall the data and the variation evident when making their final decisions about the growth of their plants. The students were often able to identify optimal growth patterns and instances where growth rates were either impeded from germination or declined after a short period of growth.

For the end-of-year questions related to Research Question 2, the recent adaptation of the SOLO model (Biggs & Collis, 1982, 1991) by Groth et al. (2021) was considered very useful in describing the longer-term learning outcomes from the activity in terms of a developmental model. When experiencing a new context and new content, as these students were, it is not surprising that at times responses failed to include concrete symbolic (CS) language and symbols in the explanation of events to answer the questions. It is important to recognise the informal visual and contextual clues students provided, indicating they were developing the intuitions required as a foundation for CS reasoning. Although the range of IK responses across the questions (Fig. 4) was from 4% (Part d) to 49% (Part b), the normative incompatible IK responses ranged from only 2% (Parts a and b) to 18% (Part c). The roughly even split of normative compatible IK and CS responses to the question, "What data did we collect to plot?" (Part b), shows an example where a focus on the specific language of what data represent is likely to support the move of students' IK responses into the CS mode. The normative incompatible IK responses to "Why is there a gap in the data?" are most likely to represent forgetfulness of some students or a continued difficulty in identifying the precise meaning of the *TinkerPlots* representation.

Although some of the representations created to show the trend in the control data were idiosyncratic by more sophisticated statistical standards (cf. Table 9), almost all students could focus on the control data and their trend upward over time. It was pleasing that over 80% of the students could write straightforward descriptions of the change in height over time in relation to the "change in the data," as well as for "comparing the control with the other treatments." Although CS responses to "what did this group learn?" fell to 68%, this may not be surprising given the new context and complex nature of the activity. The normative compatible IK responses (12%) brought the percentage responding in the context to 80%, which offers promise for future movement to the CS mode with further conversation.

Conclusion

As a culmination of the larger STEM-based project (cf. Fitzallen & Watson, 2020), the Plant Growth activity was planned to embed explicitly many of the content descriptors of the *Australian Curriculum* that can be related to make meaning in a STEM context (ACARA, 2016). The overlap across the curriculum of the critical underlying concepts and goals that are found there should reinforce efforts of teachers to make STEM connections explicit to students in the classroom. As Watson et al. (2020a) point out, many career opportunities are emerging across a range of fields linked to STEM. It is hoped that some students will continue an interest in similar topics as they continue their education.

Picking variation as the underpinning theme for the entire project reflected the obvious, but not always acknowledged, realisation that none of the STEM fields would exist except for the variation that exists across every aspect of the fields. As noted earlier, the importance of variation was not broadly recognised by the statistics education community until the 1990s, following Moore's (1990) claim that variation underpined four of his five core components of statistical thinking. This recognition has continued through the American Statistical Association's two *GAISE Reports* by Franklin et al. (2007) and Bargagliotti et al. (2020). Early work with student interviews (e.g. Watson & Kelly, 2005; Watson et al., 2003), surveys (e.g. Watson & Kelly, 2004), and classroom interventions (e.g. Lehrer & Kim, 2009; Lehrer & Schauble, 2004; Lehrer et al., 2007; Petrosino et al., 2003), has laid a foundation for studying understanding with larger interventions and meaningful contexts, providing links across the curriculum, as explored here. Hopefully, further research will be carried out, not only related to STEM in this context from Biological Science, but also more widely for topics of interest across the STEM fields.

In their analyses of students creating representations of variation, by hand, in two contexts, where the variation was of their own creation in making measurements or in the natural growth of the plants, Lehrer and Schauble (2007) emphasised the difficulties students had creating representations of the variables in the second, more complex, context. By the time the students in the study reported here reached the plant growth activity, the last activity of nine in the project, they had the support of *TinkerPlots* technology and experience using it in other activities (e.g. Fitzallen & Watson, 2020; Watson et al., 2020b). The authors speculate that the students in this study would not have been able to create reasonable representations of the data and variation in the activity by hand.

The language in the workbook questions, "within" and "between" treatments, was deliberately chosen to mirror the type of language that is used in more advanced statistics analyses when similar decisions are made about trends in experimental data. Helping students to begin thinking in such ways about data, with relevant language, can hopefully contribute to the appreciation of similar contexts encountered in later courses in the STEM fields. Similar to Watson's (2005) evidence that students' appreciation of variation precedes their appreciation of expectation, having students thinking about variation within and between variables intuitively at this level may support understanding when numerical techniques are introduced later.

In the large environment of research that is required to put STEM education firmly in the school curriculum (Honey et al., 2014), this study is a relatively small component. Although part of a larger longitudinal study, it stands on its own here, as an example of Foundational Research (p. 139), aiming to model improved learning by developing and testing methodologies and technologies (e.g. *TinkerPlots*) that will inform others. In terms of the components of an Integrated STEM Education (p. 32), the Goals and Outcomes for students are addressed here through the developing of the twenty-first-century competencies, particularly associated with integrating statistical literacy with other STEM literacies to make meaningful foundations for future growth among the fields. The study also suggests classroom design aspects to be further trialed and improved by others.

Constraints

In retrospect, it is always possible to imagine that some aspects of the plant growth activity could have been different. Allowing many choices of treatments was complex to organize and the authors believe the outcome could have been achieved with fewer choices available. Students, however, were seen to enjoy discussions in their groups on what choices to make, sometimes based on their previous out-of-school experiences with the treatments. Although the students could rarely compare progress and results directly with those in other groups of four, there was interest across the class in seeing which plants and treatments grew better.

Some of the questions on the end-of-year questionnaire could have been phrased to expect higher level SOLO responses, but overall the authors had a purpose in not doing so, given the time gap for some students and the length of the complete end-of-study questionnaire. Although some students did not respond to a few questions, generally the response rate was good. The teacher and the researchers asked and encouraged the students to complete the end-of-year questionnaire, but there was no penalty for not doing so and it did not count toward their in-class assessment by the classroom teacher. Those who completed the task appeared to enjoy thinking back on the activity.

Appendix 1

Proficiencies for Statistics the *Australian Curriculum* Years 1–6
Year 1

- **Reasoning** includes justifying representations of data and explaining patterns that have been created

Year 2

- **Reasoning** includes interpreting simple representations of data

Year 3

- **Problem-solving** includes formulating and modelling authentic situations involving planning methods of data collection and representation
- **Reasoning** includes creating and interpreting variations in the results of data collections and data displays

Year 4

- **Fluency** includes collecting and recording data
- **Reasoning** includes communicating information using graphical displays and evaluating the appropriateness of different displays

Year 5

- **Reasoning** includes posing appropriate questions for data investigations and interpreting data sets.

Year 6

- **Problem-solving** includes interpreting secondary data displays

Appendix 2

Treatment concentrations for the Plant Growth activity.

	Salt	**Sugar**	**Vinegar**	**Liquid fertiliser**	**Soil type***
Treatment 1	1 g/L 1 tsp / 5 L (0.1%)	10 g/L 2 tsp/L (1%)	1 ml/L 1 tsp / 5 L (0.1%)	0.3 ml/L 1 tsp / 15 L (0.03%)	Worm castings
Treatment 2	1.5 g/L 1½ tsp / 5 L (0.15%)	50 g/L 10 tsp/L (5%)	10 ml/L 10 tsp /5 L (1%)	1.5 ml/L 5 tsp / 15 L (0.15%)	Potting mix without fertiliser
Treatment 3	10 g/L 10 tsp / 5 L (1%)	100 g/L 20 tsp/L (10%)	20 ml/L 20 tsp / 5 L (2%)	3 ml/L 10 tsp / 15 L (0.3%)	Organic compost
Control	Water	Water	Water	Water	Vermiculite

Student checklist for the Plant Growth activity.

Tick the type of plant your group will use in your experiment.

Plant Type: Wheat ☐ Radish ☐

Tick the treatment condition your group has decided on.

Treatment: Salt ☐ Sugar ☐

Vinegar ☐ Liquid Fertilizer ☐

Soil type ☐

Tick the type of soil your group will use. If you choose "soil type" your group will use a different type of soil in each bottle.

Soil Type: Vermiculite ☐ Compost ☐

Potting mix ☐ Worm castings ☐

Write in the table below the name of the person responsible for each of the treatments:

Control	Treatment 1	Treatment 2	Treatment 3

Example of data collection table for each time measurements were made.

Today's date is: Friday 18th May, 2018

Number of days since seeds were planted: 7

Treatment type:

Student's Name	**Bottle**	**Height of the 4 tallest plants in each bottle**			
		Plant 1	**Plant 2**	**Plant 3**	**Plant 4**
	Control				
	Treatment 1				
	Treatment 2				
	Treatment 3				

Acknowledgements Watson, Fitzallen, and Kelly acknowledge the assistance of Dr. Caroline Smith, who supported the implementation of the activity, and express extreme sadness at the untimely passing of their co-author Suzie Wright.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions. This study was funded as part of an Australian Research Council (ARC) Discovery Grant (DP150100120). Any opinions, findings, conclusions, or recommendations reported are those of the authors and do not necessarily reflect the views of the ARC.

Declarations

Ethical approval The longitudinal study for which the Plant Growth activity was a part had ethics approval from the Tasmanian Social Sciences Human Research Ethics Committee (approval number H0015039).

Informed consent The 64 students whose data were used in the study had written parental permission for their data to be used. Similarly, the parents of the girl in Fig. 1 gave written permission to use her photo. The photos used were taken by co-author Suzie Wright.

Conflict of interest The authors declare no competing interests.

References

Allmond, S., & Makar, K. (2014). From hat plots to box plots in *TinkerPlots*: Supporting students to write conclusions which account for variability in data. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18). International Statistical Institute. http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_2E1_ALLMOND.pdf

Anderson, J., & Li, Y. (2020). *Integrated approaches to STEM education: An international perspective.* Springer. https://doi.org/10.1007/978-3-030-52229-2

Australian Academy of Science. (2016). *Primary connections*: *Rising salt Year 6 Biological sciences*. https://primaryconnections.org.au/sites/default/files/unit-file-downloads/Risingsalt.pdf

Australian Curriculum, Assessment and Reporting Authority. (2016). *ACARA STEM Connections Project Report*, June, 2016. https://www.australiancurriculum.edu.au/media/3220/stem-connections-report.pdf

Australian Curriculum, Assessment and Reporting Authority. (2019). *Australian Curriculum.* ACARA. https://www.australiancurriculum.edu.au/

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. A. (2020). *Pre-K-12 Guidelines for Assessment and Instruction in Statistics Education II* (GAISE II). American Statistical Association. https://www.amstat.org/asa/files/pdfs/GAISE/GAISEIIPreK-12_Full.pdf

Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press.

Biggs, J. B., & Collis, K. F. (1991). Multimodal learning and the quality of intelligent behaviour. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 57–76). Lawrence Erlbaum.

Chick, H. L. (2007). Teaching and learning by example. In J. Watson & K. Beswick (Eds.), *Mathematics: Essential research, essential practice. Proceedings of the 30th Annual Conference of the Mathematics Education Research Group of Australasia* (Vol. 1, pp. 3–21). MERGA.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *American Mathematical Monthly, 104*(9), 801–823. https://doi.org/10.2307/2975286

Creswell, J. W. (2013). *Research design: Quantitative, qualitative and mixed method approaches* (2nd ed.). SAGE Publications.

Delen, I., Morales, C. J., & Krajcik, J. (2020). Missing coherence in STEM education: Creating design-based pedagogical content knowledge in a teacher education program. In J. Anderson & Y. Li (Eds.), *Integrated approaches to STEM education: An international perspective* (pp. 361–383). US: Springer. https://doi.org/10.1007/978-3-030-52229-2_20

Early Childhood STEM Working Group. (2017). *Early STEM matters: Providing high-quality STEM experiences for all young learners.* UChicago STEM Education and Erikson Institute.

Engler, J. (2012). STEM education is the key to the U.S.'s economic future. *U.S. News and World Report Civic.* https://www.usnews.com/opinion/articles/2012/06/15/stem-education-is-the-key-to-the-uss-economic-future

Fielding-Wells, J., & Hillman, J. (2018). Supporting young students emerging understandings of centre through modelling. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward.* (Proceedings of the Tenth International Conference on Teaching Statistics July, 2018, Kyoto, Japan). International Statistical Institute. https://iaseweb.org/icots/10/proceedings/pdfs/ICOTS10_2B1.pdf?1531364242

Fitzallen, N., & Watson, J. (2020). Using the practice of statistics to design students' experiences in STEM. In B. Shelley, K. te Riele, N. Brown, & T. Crellin (Eds.), *Harnessing the transformative power of education* (pp. 74–99). Koninklijke Brill.

Fitzallen, N., Watson, J., & Wright, S. (2017). The heat is on! Using a stylised graph to engender understanding. *Australian Primary Mathematics Classroom, 22*(2), 3–7.

Fitzallen, N., Wright, S., & Watson, J. (2019). Focusing on data: Year 5 students making STEM connections. *Journal of Research in STEM Education, 5*(1), 1–19. https://doi.org/10.51355/jstem.2019.60

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework*. American Statistical Association. http://www.amstat.org/education/gaise/

Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Lawrence Erlbaum.

Green, D. (1993). Data analysis: What research do we need? In L. Pereira-Mendoza (Ed.), *Introducing data analysis in the schools: Who should teach it?* (pp. 219–239). International Statistical Institute.

Groth, R. E., Austin, J. W., Naumann, M., & Rickards, M. (2021). Toward a theoretical structure to characterize early probabilistic thinking. *Mathematics Education Research Journal, 33*(2), 241–261. https://doi.org/10.1007/s13394-019-00287-w

Harland, D. J. (2011). *STEM student research handbook*. National Science Teachers Association.

Holmes, P. (1980). *Teaching Statistics 11–16*. Schools Council Publications and W. Foulsham & Co.

Honey, M., Pearson, G., & Schweingruber, H. (Eds.). (2014). *STEM integration in K-12 education: Status, prospects, and an agenda for research*. Academy of Sciences.

James, G., & James, R.C. (Eds.). (1959). *Mathematics Dictionary*. D. Van Nostrand Company.

Kazak, S., Fujita, T., & Wegerif, R. (2014). Year 6 students' reasoning about random 'bunny hops' through the use of *TinkerPlots* and peer-to-peer dialogic interactions. In K. Makar, B. deSousa, & R. Gould (Eds.), *Sustainability in statistics education* (Proceedings of the 9th International Conference on the Teaching of Statistics, Flagstaff, Arizona, July 13–18). International Statistical Institute. http://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_9F1_KAZAK.pdf

Khairiree, K., & Kurusatian, P. (2009). Enhancing students' understanding of statistics with TinkerPlots: Problem-based learning approach. In W.-C. Yang, M. Majewski, T. deAlwis, & Y. Cao (Eds.), *Electronic Proceedings of the Fourteenth Asian Technology Conference in Mathematics, Beijing, China.* https://atcm.mathandtech.org/ep2009/papers_full/2812009_17324.pdf

Kirkpatrick, E. M. (Ed.). (1983). *Chambers 20th Century Dictionary* (3rd ed.). Chambers.

Kitchens, L. J. (1998). *Exploring statistics: A modern introduction to data analysis and inference*. Brooks/Cole.

Konold, C. (2007). Designing a data analysis tool for learners. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 267–291). Lawrence Erlbaum.

Konold, C., Harradine, A., & Kazak, S. (2007). Understanding distributions by modeling them. *International Journal of Computing and Mathematics Learning, 12*, 217–230.

Konold, C., & Lehrer, R. (2008). Technology and mathematics education: An essay in honor of Jim Kaput. In L. English (Ed.), *Handbook of international research in mathematics education* (2nd ed., pp. 49–72). Routledge.

Konold, C., & Miller, C. D. (2015). *TinkerPlots: Dynamic data exploration* [Computer software, Version 2.3.2]. Learn Troop. https://www.tinkerplots.com/

Lehrer, R., & Kim, M. (2009). Structuring variability by negotiating its measure. *Mathematics Education Research Journal, 21*(2), 116–133.

Lehrer, R., Kim, M., & Schauble, L. (2007). Supporting the development of conceptions of statistics by engaging students in measuring and modeling variability. *International Journal of Computers for Mathematical Learning, 12*, 195–216.

Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal, 41*(3), 635–680.

Lehrer, R., & Schauble, L. (2007). Contrasting emerging conceptions of distribution in contexts of error and natural variation. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 149–176). Lawrence Erlbaum Associates.

Lovett, M. C., & Shah, P. (Eds.). (2007). *Thinking with data*. Lawrence Erlbaum.

Lyden, S., Ward, L., Fitzallen, N., & Panton, L. (2018). Exploring a STEM education pedagogy: Teachers' perceptions of the benefits of an extended integrative STEM learning program. In *Integrated education for the real world* (5th International STEM in Education Conference, Brisbane, pp. 422–429).

Mackenzie, N., & Knipe, S. (2006). Research dilemmas: Paradigms, methods and methodology. *Issues in Educational Research, 16*(2), 193–205.

Mayes, R. (2019). Quantitative reasoning and its role in interdisciplinarity. In B. Doig, J. Williams, D. Swanson, R. Borromeo Ferri, & P. Drake (Eds.), *Interdisciplinary mathematics education: The state of the art and beyond* (pp. 113–133). Springer.

Mildenhall, P., Cowie, B., & Sherriff, B. (2019). A STEM extended learning project to raise awareness of social justice in a year 3 primary classroom. *International Journal of Science Education, 41*(4), 471–489. https://doi.org/10.1080/09500693.2018.1560514

Moore, D. S. (1990). Uncertainty. In L. A. Steen (Ed.), *On the shoulders of giants: New approaches to numeracy* (pp. 95–137). National Academy Press.

National Research Council. (2013). *Next generation science standards: For states, by states.* The National Academies Press.

Office of the Chief Scientist. (2013). *Science, technology, engineering and mathematics in the national interest: A strategic approach*. Commonwealth of Australia. ISBN 978-1-921916-83-0

Pecen, R., Humston, J. L., & Yildiz, F. (2012). Promoting STEM to young students by renewable energy applications. *Journal of STEM Education, 13*(3), 62–72.

Petrosino, A. J., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*(2–3), 131–156. https://doi.org/10.1080/10986065.2003.9679997

Shaughnessy, J. M. (1997). Missed opportunities in research on the teaching and learning of data and chance. In F. Biddulph & K. Carr (Eds.), *People in mathematics education* (Vol. 1, pp. 6–22). MERGA Proceedings of the 20th annual conference of the Mathematics Education Research Group of Australasia.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook on research on mathematics teaching and learning* (pp. 957–1009). Information Age Publishing.

Smith, C., Fitzallen, N., Watson, J., & Wright, S. (2019). The practice of statistics for STEM: Primary students and pre-service primary teachers exploring variation in seed dispersal. *Teaching Science, 65*(1), 38–47.

Smith, E. L., & Anderson, C. W. (1984). Plants as producers: A case of elementary science teaching. *Journal of Research in Science Teaching, 21*(7), 685–698.

Steffensen, L. (2020). Climate change and students' critical competencies: A Norwegian study. In J. Anderson & Y. Li (Eds.), *Integrated approaches to STEM education: An international perspective* (pp. 271–293). Springer. https://doi.org/10.1007/978-3-030-52229-2_15

Stohlmann, M., Moore, T. J., & Roehrig, G. H. (2012). Considerations for teaching integrated STEM education. *Journal of Pre-College Engineering Education Research, 2*(1), 4. https://doi.org/10.5703/1288284314653

Tytler, R., Mulligan, J., Prain, V., White, P., Xu, L., Kirk, M., Nielsen, C., & Speldewinde, C. (2021). An interdisciplinary approach to primary mathematics and science learning. *International Journal of Science Education, 43*(12), 1926–1949. https://doi.org/10.1080/09500693.2021.1946727

Tytler, R., Williams, G., Hobbs, L., & Anderson J. (2019). Challenges and opportunities for a STEM interdisciplinary agenda. In B. Doig, J. Williams, D. Swanson, R. Borromeo Ferri, & P. Drake (Eds.), *Interdisciplinary mathematics education: The state of the art and beyond* (pp. 51–81). Springer.

Wang, S., Ching, Y., Yang, D., Swanson, S., Baek, Y., & Chittorri, B. (2020). Developing US elementary students' STEM practices and concepts in an afterschool integrated STEM project. In J. Anderson & Y. Li (Eds.), *Integrated approaches to STEM education: An international perspective* (pp. 205–226). US: Springer. https://doi.org/10.1007/978-3-030-52229-2_12

Watson, J. M. (2005). Variation and expectation as foundations for the chance and data curriculum. In P. Clarkson, A. Downton, D. Gronn, M. Horne, A. McDonough, R. Pierce, & A. Roche (Eds.), *Building connections: Theory, research and practice* (pp. 35–42). MERGA Proceedings of the 28th annual conference of the Mathematics Education Research Group of Australasia, Melbourne.

Watson, J. M. (2009). The influence of variation and expectation on the developing awareness of distribution. *Statistics Education Research Journal, 8*(1), 32–61.

Watson, J., & Fitzallen, N. (2016). Statistical software and mathematics education: Affordances for learning. In L. English & D. Kirshner (Eds.), *Handbook of international research in mathematics education* (3rd ed., pp. 563–594). Taylor and Francis.

Watson, J., & Fitzallen, N. (2021). What sense do children make of "data" by Year 3? In Y. H. Leong, B. Kaur, B. H. Choy, J. B. W. Yeo, & S. L. Chin (Eds.), *Excellence in mathematics education: Foundations and pathways* (pp. 409–416). MERGA Proceedings of the 43rd annual conference of the Mathematics Education Research Group of Australasia.

Watson, J., Fitzallen, N., & Chick, H. (2020a). What is the role of statistics in integrating STEM education? In J. Anderson & Y. Li (Eds.), *Integrated approaches to STEM education: An international perspective* (pp. 91–116). Springer. https://doi.org/10.1007/978-3-030-52229-2_6

Watson, J., Fitzallen, N., English, L., & Wright, S. (2020b). Introducing statistical variation in year 3 in a STEM context: Manufacturing licorice. *International Journal of Mathematical Education in Science and Technology, 51*(3), 354–387. https://doi.org/10.1080/0020739X.2018.1562117

Watson, J., Fitzallen, N., Fielding-Wells, J., & Madden, S. (2018). The practice of statistics. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 105–137). Springer.

Watson, J., Fitzallen, N., & Kelly, B. (2022a). *Interpolation and extrapolation in Grade 5: Exploring the data without formal mathematics*. Manuscript under review.

Watson, J., Fitzallen, N., Wright, S., & Kelly, B. (2022b). Characterizing student experience of variation within a STEM context: Improving catapults. *Statistics Education Research Journal, 21*(1), 11. https://doi.org/10.52041/serj.v21i1.7

Watson, J.M., & Kelly, B. A. (2004). Statistical variation in a chance setting: A two-year study. *Educational Studies in Mathematics, 57*(1), 121–144. https://www.jstor.org/stable/4150317

Watson, J. M., & Kelly, B. A. (2005). The winds are variable: Student intuitions about variation. *School Science and Mathematics, 105*(5), 252–269.

Watson, J. M., Kelly, B. A., Callingham, R. A., & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *International Journal of Mathematical Education in Science and Technology, 34*(1), 1–29. https://doi.org/10.1080/0020739021000018791

Watson, J. M., & Moritz, J. B. (1998). Longitudinal development of chance measurement. *Mathematics Education Research Journal, 10*(2), 103–127.

Watson, J. M., & Moritz, J. B. (2000). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior, 19*, 109–136.

Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation, 37*, 3–14.

Wood, A. J., & Roper, J. (2000). A simple & nondestructive technique for measuring plant growth & development. *The American Biology Teacher, 62*(3), 215–217. https://www.jstor.org/stable/4450877

Wright, S., Watson, J., Smith, C., & Fitzallen, N. (2021). What makes plants grow best? Using a data-based STEM approach to find out. *Teaching Science, 67*(2), 37–50.