SCAFFOLDING YEAR 8 STUDENTS' STATISTICAL MODELLING REASONING USING FOLLOW UP TASKS TO A MODEL ELICITING ACTIVITY

Anne Patel and Maxine Pfannkuch
Department of Statistics, The University of Auckland, New Zealand
ablundellnz@gmail.com

In response to the problem about how to build on students' initial conceptions shown in Model Eliciting Activities (MEAs), recent mathematics education research proposed using Follow Up Tasks (FUTs) within the same context as the MEA to scaffold and consolidate students' reasoning. In this paper, we focus on four 12-year-old students who had completed a schoolbag weight MEA using TinkerPlots. Based on a qualitative analysis of their reasoning and artefacts produced during the MEA, FUTs were designed to introduce them to notions such as representative samples, assessing model fit and conditioning. Findings suggest that FUTs have the potential to develop in students more sophisticated statistical modelling conceptions and competencies, including the ability to adapt their MEA schoolbag weight model to incorporate conditioning.

INTRODUCTION

Some researchers (e.g., Lehrer, 2015) advocate the benefits of a modelling approach for developing inferential reasoning as it is more aligned with how novice students think. A modelling approach has been made possible by *TinkerPlots* (Konold & Miller, 2011) as it enables students to create models to produce simulated data using random generating devices. Some tools in the *TinkerPlots* model building environment are: spinners, balls in urns and continuous distribution curves which can be linked to build models. A modelling approach, furthermore, reconnects the chance and data strands of the curriculum (Konold & Kazak, 2008). For example, conditioning in the statistical sense is traditionally taught through a mathematical lens using tables, formula or tree diagrams, thus losing the connection to the provenance of data.

The premise of model construction is to explore phenomena arising in data. Students must consider multiple factors and their distributions that underpin the random variation seen in data. The simulated data produced by the model can be fashioned to match features of the real data distribution. This fashioning involves proportional and aggregate reasoning about distributions of factors that contribute to the model. Once an acceptable model fit is obtained, the model can be used to ask and answer "what if" questions, to predict what alterations to the model may mean in the context of the problem. The ideal statistical modelling task is grounded in an *authentic* problem solving context, which serves to strengthen students' reasoning when mapping aspects of reality to the model world and vice-versa, as well as providing motivation to arrive at a solution to the problem through the modelling process (Maaß, 2006; Dierdorp, Bakker, Eijkelhof, & van Maanen, 2011). Research (e.g., Pfannkuch, 2011, however, shows that context can both support and distract from developing statistical inference notions, and therefore both task design and context are critical in supporting students' reasoning through the modelling process.

In mathematics education research there is renewed interest in models and modelling. Mathematics education researchers such as Lesh and Doerr (2003) advocate learning sequences that begin with a Model Eliciting Activity (MEA) that elicits knowledge students bring to the task. During an MEA a myriad of models are potentially developed by a class of students. Therefore, the development of students' reasoning about *effective* and *efficient* models as well as statistical concepts such as inference becomes an issue for the learning sequences that follow an MEA. Lesh and Doerr (2003) suggest Follow Up Tasks (FUTs) that involve altering and extending models. However, they do not suggest what these tasks might look like. Yoon, Patel, Sullivan and Radonich (2011) took up the challenge of building on and developing students' mathematical, modelling and model communication competencies by designing FUTs based in the same context as the MEA. They sourced ideas and inspiration directly from students' models developed during an MEA. For example, in a proportional reasoning MEA, students either used an additive model or a multiplicative model. FUTs were designed to shift the students using an additive model to the more powerful multiplicative model. These FUTs provided students with an opportunity to clarify and consolidate their thinking or see the necessity of adopting a new model. Yoon et al.'s study

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

provided blue prints for FUT design, which drew on Vygotsky's Zone of Proximal Development learning theory, whereby FUTs consisting of a mixture of written and oral responses of students elicited from the MEA are led by a "knowledgeable other". FUTs can also present contrasting ideas to be debated between students and can make use of pre-built models to explore the structure of models or to do further model exploration by adapting or extending models (Pratt, 2011).

This research sets out to consolidate and scaffold twelve year-old students' inferential reasoning using FUTs by continuation of a problem situation explored in an MEA. As MEAs are grounded in constructivist views of education, they place an emphasis on contexts that are well understood by students while also providing motivation to solve a problem. Our FUTs aim to capitalise on students' contextual knowledge and modelling experiences uncovered in the MEA and integrate contextual and statistical reasoning at a much deeper level. Hence the research question addressed in this paper is: How do our FUTs, set in the same context as an MEA, support the development of novice students' statistical reasoning about representativeness, model fit and uncertainty, and conditioning?

BACKGROUND TO RESEARCH AND PARTICIPANTS

In the first year of this study, set in a Year 7 and 8 Intermediate school, six eleven-year-old students, working in pairs in two-hour sessions over a six-week period, were introduced to the *TinkerPlots* data exploration and sampler model building components. During the first lesson of the learning sequence the class the six students were in completed a CensusAtSchool (CAS) survey, an item of which was to record the weight of their school bags. The learning sequence culminated in an MEA that presented a problem about students carrying overweight school bags. The MEA used a media article reporting on a Massey University study that quantified the percentage of students with overweight bags in New Zealand and suggested that school bags should not weigh more than 10% of a student's body weight. Hence, for the MEA, the students used their class's school bag weights as the real data distribution and modelled the situation in *Tinkerplots* by conducting investigations into the distributions of weights of items found in their school bags. They then used their models to solve the problem of finding the number of students in their school with overweight school bags and to answer a "what if" question about the impact of becoming a Bring Your Own Device school. The students then communicated their findings in a letter to their Principal.

During the MEA students' verbalisations and interactions with the *TinkerPlots* environment were captured using Camtasia. A micro-analysis of two students' interactions resulted in a draft theoretical Statistical Modelling framework. This framework, which was based on the Mathematical Modelling framework of Stillman, Galbraith, Brown and Edwards (2007), characterised these two students' reasoning, including the statistical and modelling notions that needed further support. Armed with this knowledge, FUTs were designed to focus on developing their notions of representativeness, model fit and uncertainty to a more sophisticated level. For model fit and uncertainty we drew on the work of Konold and Kazak (2008) about how to focus students' attention on different aspects of distribution such as centre, range, shape, variation and the effect of sample size. The FUTs were deliberately set in the context of school bag weights and hence a task was designed to extend the model students had developed in the MEA to include notions of conditioning. As the students had experience of separating out plots of data based on factors such as gender, two FUTs were designed to facilitate students exploring factors and situations involved in overweight school bags. These required students to investigate and then condition bag weights on factors such as Year level and gender.

In the second year of this study, the same six students who were now in Year 8, trialled the FUTs. The length and complexity of the modelling process involves a large investment of time spent investigating the context and problem situation. Rather than changing context after the MEA, the FUTs were designed to utilise students' contextual knowledge and understanding about the problem situation developed during the MEA as they seemed more likely to support and build on students' naïve statistical understandings. The data gathered during the trials of the FUTs included Camtasia footage and TinkerPlots artifacts, a description of which is now presented.

DESCRIPTION OF STUDENTS' REASONING PROCESSES DURING THE FUTS

Four students' reasoning processes for the FUTs that supported their ideas about representativeness, model fit and uncertainty, and conditioning are discussed by indicating what the MEA revealed about students' reasoning, describing the FUT and then reporting on the students' reactions.

Representativeness

For the MEA the students based their model construction on bag weights from their class, which was the only data available. At the time they did not question the use of these data, but several students voiced their concerns that their class data may not accurately reflect the bag weights of the whole school. To encourage students to explore the concept of using a representative sample, the FUTs were designed to focus students' attention on differences between their class data and a random sample of Year 7 and 8 students' bag weights taken from the CAS database. An FUT presented the plot of a random sample of bag weights (n=250) that had a skewed distribution. Students were asked to compare this with a plot of their class data, which was symmetric. They identified and described differences in the ranges and shapes and noted the similarity of the means. The students seemed simultaneously aware of the variation within and yet homogenous nature of the bag weights of students in their class. They used contextual reasoning for why students' bag weights in the random sample were positively skewed, such as, it included "both Year 7s and Year 8s". They suggested diverse activities that other students and classes may participate in that they did not, which they theorised would have an impact on bag weights. They also commented on the large proportion of students in the random sample who had no bag or very light bags. This was not the case in their class, as every student in the class had a school bag, with the minimum bag weight being 1.5kgs. Lee noted a possible reason for the lighter bag weights in the random sample was "probably the school because they could be BYOD (Bring Your Own Device) or not, or, could get lunch from school [and] not have to carry it." In reasoning about the upper tail in the random sample, Dan theorised this could be due to students not telling the truth, as they queried the 17kg maximum bag weight. They concluded that the student could have lied or there was a small chance that the datum was correct if a student carried two bags, for example, a sports bag and a school bag.

Another FUT presented opposing views about the representativeness of data that the students' models should be based on.

- Student A: "We should use the bag weights of our class as the real data because we know our class did the census [the CAS survey] properly, so the data will be accurate."
- Student B: "But a random sample will mean we get a better understanding of the bag weights in the whole school, our class is only Year 7 students and we need students from other classes and Year 8s to get a better picture of all the bag weights in the school, not just our class."

After discussing these two claims, the students agreed with Student B that using a random sample was better than using only their class data as a basis for building a model. They realised that their class data had not been the most reliable representation of the bag weights for the whole school population. This was an important lesson in model construction about the need for the real data to be representative of the population being modelled.

Model Fit and Uncertainty

In the MEA, when students were matching the real data to the simulated data, they focussed only on matching the ranges and the percentage of overweight bags. They did not appear to match the centre or the shape of the distributions, nor did they use tools to enumerate the central tendency of the simulated distribution. The students were aware of the variation in the ranges and the percentage of overweight bags when they ran multiple simulations, but they did not seem to notice the variation in the centres or shape or notice the reduced variation in the range or average when they increased the number of trials. Furthermore, when they communicated the percentage of overweight bags they only gave one value and not an interval of plausible values based on many simulations. Therefore, FUTs were designed to focus students' attention on matching three distribution aspects–centre, shape and range–from multiple distributions in order to judge model fit and then looking to find a range of percentages of overweight school bags to describe the uncertainty suggested by the model.

In order to draw their attention to the variation in the centre, range and shape occurring in the simulated distribution, they were given real data from a random sample and four simulated distributions for n=35 and n=100, then asked to evaluate the model fit as Bad/Ok/Great based on the three aspects. The students read the average and range easily from the distributions and sketched the shape of the distribution by closely following the outline of the plots. They then amalgamated this knowledge to judge the model fit. There was disparity between students in their model fit judgements ranging from a bad to a great model fit. Some students changed their minds from Great to Ok or Bad, after group discussions. The FUT also drew their attention to the sample size effect, by getting students to write the averages for the four simulated distributions for each sample size and asking what they noticed about the variation in the averages.

To promote the idea that values from one simulation should not be used, rather, a good model fit needs to be based on a fit that occurred *most of the time*, the students were given a scenario of six simulated bag weight distributions (n=100).

- Student C: "use the second plot because the average of the simulation was the same as the real data"
- Student D: "We need to get the model giving a good fit to the real data **most of the time**, not just once! All of the averages apart from that one *[Plot 2]* are less than 3.5kgs. I think the average of the simulated data should be moving around on both sides of the real average, not always to one side of it."

The students had to state who they agreed with and why. After discussion, there was general agreement that multiple plots needed to be considered to inform judgement about model fit. Also, they agreed that these plots may not produce the exact value, for say the average, but should be *either side* of the average, *most of the time*.

The next FUT used a pre-built bag weight model, which required students to run multiple simulations to judge and justify model fit by recording the intervals for the simulated average and range and comparing these with the real average and range. They judged the pre-built model fit as Bad or Ok. The FUT, in an attempt to raise awareness about critiquing models, suggested two possibilities to improve the model: (1) separate factors for food and water and (2) include a pencil case/equipment factor. When asked if they agreed with the first critique of the model Lee wrote: "Yes, because some people bring heaps of food and no water or the other way around. This can easily change our data". For the second critique, the students generally disagreed. Lee writing: "No, because most year 7 and 8s leave their pencil cases in their desks and most pencil cases are really light." They then altered and refined the model based on the first critique, and ran it multiple times comparing the intervals for the average and range against the real data, until they produced a Great model fit. The next part of the FUT required students to give an *interval* for the predicted percentage of underweight bags. This excerpt shows the students looking at the variation in the percentage of underweight bags occurring in multiple simulations and deciding on the interval:

Ali: Run it a few times. Ok, between 45 and 47 [%, names interval for two simulations], 45 and 53, 44 and 42, 55 and 46.

Lee: How about we go between 43 and 55?

Ali: Yep ok. No let's go 40 to 55 [believes 40 is a potential value].

Lee: Are you sure?

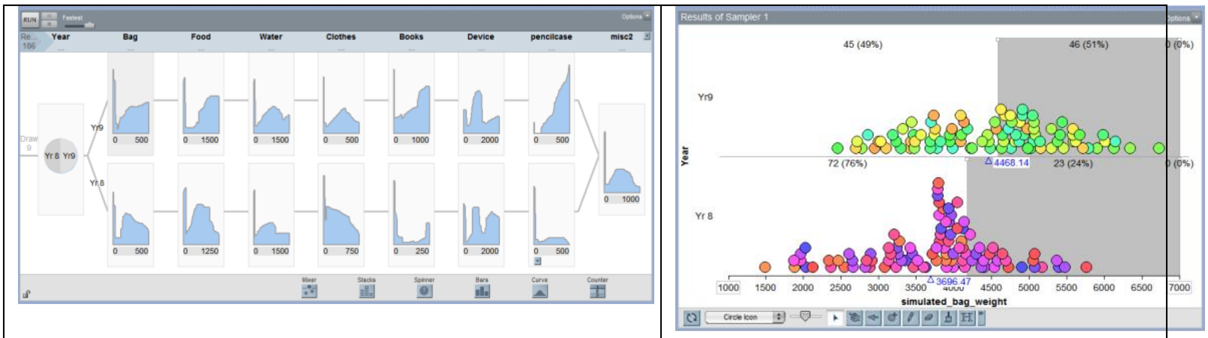
Ali: Cos don't forget cos that's 43 [percentage underweight for real data].

The students' new focus on the amount and range of variation in the percentage of underweight bags illustrates their growing awareness of the necessity to communicate a range of values rather than a single value. Focusing on variation in the average and predicted percentages served to establish initial notions about uncertainty and the need to communicate a neighbourhood of values (Lehrer, 2015).

Conditioning

Because students had only used a linear modelling process in the learning sequence, FUTs were designed that required them to adapt their bag weight models using the new notion of conditioning. The first FUT required students to investigate and model the difference between girls' and boys' bag weights. The students were required to investigate a *random* sample from the CAS database from the Otago region in New Zealand. Nico and Dan initially built a curve tool for bag weight as the first factor but after adding several more factors, they realised gender had to be included earlier in the model as other factors depended on the outcome of gender. They realised gender assignment required a spinner tool to represent the two outcomes, rather than a curve tool. They chose to differentiate the factors of clothes, food, device and water for girls and boys. They did this by keeping the range of each factor the same and then altering the shape of the distributions according to gender. They then merged their model, with assistance from the researcher and created two factors they theorised would be the same for boys and girls called “miscellaneous” and “bag2” (the weight of an empty bag). Ali and Lee conditioned their model not only on the students’ gender but also on whether the student carried sports gear or not. This new aspect to model building appeared to provide interesting avenues of exploration for these students. Ali and Lee also kept the range of the factors the same, whilst changing the distributions.

The second FUT required the students to examine the problem of Year 8 Intermediate students, who had desks to store their school equipment, moving to a new secondary school environment in Year 9, where students were expected to carry all their equipment in their bags at all times. The Year 8 students were anecdotally aware of Year 9 students beginning secondary school with larger and heavier bags. For this task Ali and Lee not only changed the shape of the distributions but also differentiated the ranges of factors for each year level. When building their model (Fig. 1a) conditioned on year level, Nico and Dan noted, “this (Year level) is our gender”, comparing the conditioning factor in the previous FUT. An important feature of the session was for students to view the other groups’ models and discuss similarities and differences.



(a) Model constructed for bag weights for each Year level | (b) Resultant simulated plots

Figure 1. Nico and Dan’s model for bag weights conditioned on Year 8 and 9 levels

When asked about what the “hills and valleys” represented in their distribution for device (Fig. 1a), Nico replied “that is the frequency or chance of that weight happening”. He stated the first hill on zero was “no device”, the second hill was “phones or ipads”, and the third hill was “computers or laptops” and gave possible combinations of these devices that contribute to bag weights. Once the plots of year level versus bag weight were created (Fig. 1b), both pairs of students firstly compared averages for Year 8 and 9 students’ bag weights and refined the models to get a close match to the average for the real data. They did this by altering both the shape of the distribution and the ranges of factors based on contextual reasoning such as “All the different gear you need (in Year 9) and also your lunch might change, you might start buying lunch at school.” The placement of the conditioning factor is an important aspect of model construction. While the factor the conditioning is based on can be placed anywhere in the model, the factors that depend on the conditioning factor must be placed after this factor. Thus, the structure of the model must be reasoned about, as cause and effect become a consideration, which these students demonstrated.

CONCLUSION AND DISCUSSION

Being immersed in the context of school bag weights resulted in drawing the attention of these novice modellers to the notion of representativeness, to aspects of distributions to judge model fit in the presence of uncertainty, and to conditioning simulated data based on underlying factors that influence phenomenon seen in real data. The use of pre-built models that resembled models built by students in the MEA meant that if students did not succeed in building a model that produced an OK model fit, which is noted as a concern of MEAs, they can still be supported to use and adapt models set in the same context as the MEA. Through FUTs students can be made aware of different structures of models and of the reasoning, assumptions and limitations behind models, thus supporting how students' understanding of the real-world situation is mapped onto chance-based models. These students also became aware of the stability that accompanies larger sample sizes resulting in narrower ranges for predictions when using their models. The dynamic nature of conditioned models and the ease with which large numbers of trials can be generated in *TinkerPlots* provided a new visual and physical element to students' conditional reasoning. Moreover, the conditioning tasks required comparing two groups of simulated data with their respective real data distributions. This act appeared to focus students' attention on features such as average and range, because these differentiated the two groups, thus supporting students to shift to an aggregate understanding of distribution (cf. Konold & Kazak, 2008).

We hypothesise that using FUTs that saturate students in a single context that they are personally familiar with seemed to help these students to develop some more statistical inference notions through a modelling approach. We also conjecture that using the same context as the MEA reduced the cognitive load on the students and the complexity of the modelling process and helped to encourage the shuttling back and forth between the model world and the real world. These findings seemed to confirm Yoon et al.'s (2011) contention that FUTs help support the development of students' reasoning. These novice modellers were also exposed to language, concepts and processes that underpin statistical inference, and although they do not have a full understanding of all of the concepts explored, they made progress towards developing inferential ideas through an active modelling approach (Lehrer, 2015). Perhaps most importantly, the students had an experience where they were involved in a problem that affects the everyday context of their lives in a meaningful way (Dierdorp et al., 2011) and they were exposed to how the modelling process, based on chance and data, encapsulates a problem and suggests actions and decisions.

REFERENCES

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011). Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning*, *13*, 132–151.

Konold, C. & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, *2*(1). http://repositories.cdlib.org/uclastat/cts/tise/vol2/iss1/art1.

Konold, C., & Miller, C. (2011). *TinkerPlots™* Version 2.3 [computer software]. Amherst, MA: Learn Troop.

Lehrer, R. (2015). Developing practices of model-based informal inference. In *Proceedings of the Ninth International Research Forum on Statistical Reasoning, Thinking, and Literacy (SRTL9, 26 July – 3 August, 2015)* (pp. 76–86). Paderborn, Germany: University of Paderborn.

Lesh, R., & Doerr, H. M. (Eds.). (2003). *Beyond constructivism: Models and modeling perspectives on mathematics problem solving, learning, and teaching.* Mahwah, NJ: Lawrence Erlbaum Associates.

Maaß, K. (2006). What are modeling competencies? *ZDM – The International Journal on Mathematics Education*, *38*(2), 113–142.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning*, 13(1&2), 27–46.

Pratt, D. (2011). Re-connecting probability and reasoning from data in secondary school teaching. *Proceedings of the 58th International Statistical Institute World Statistical Congress, Dublin*, (pp. 890–899). The Hague, The Netherlands: International Statistical Institute.

Stillman, G. Galbraith, P., Brown, J., & Edwards, I. (2007). A framework for success in implementing mathematical modeling in the secondary classroom. In J. Watson & K. Beswick (Eds.), *Proceedings of the 30th annual conference of the Mathematics Education Research Group of Australasia*, (Vol. 2, pp. 688–697). Adelaide, Australia: MERGA Inc.

Yoon, C., Patel, A., Radonich, P., & Sullivan, N. (2011). Learning environments with mathematical modeling activities. *Teacher Learning Research Initiative*, Wellington, New Zealand: New Zealand Council for Educational Research. Retrieved from: http://www.tlri.org.nz/sites/default/files/projects/9274_finalreport.pdf