The 15th International Congress on Mathematical Education
Sydney, 7-14 July, 2024

TOWARDS A FRAMEWORK FOR ASSESSING INTEGRATED STATISTICAL AND COMPUTATIONAL THINKING

Anna Fergusson and Maxine Pfannkuch
University of Auckland, New Zealand

Data science involves the use of modern data and related technologies, necessitating the use of computational tools and associated modelling approaches. Consequently, a key component of data science education is the teaching of integrated statistical and computational thinking. In this paper, we present a characterisation of integrated statistical and computational thinking in terms of shuttling between or connecting the contextual, statistical, and computational spheres. Based on observations of learners' thinking practices while they engaged with statistical modelling tasks that introduced code-driven tools, we developed a hypothesised framework that identifies eight potential observable integrated statistical and computational thinking practices. The framework may provide useful support to data science teachers with respect to creating learning tasks.

INTRODUCTION

A common thread to discussions about data science education is that learners need to integrate both statistical and computational thinking (e.g., De Veaux et al., 2017) to support them to understand and interpret modern data (Nolan & Temple Lang, 2010; Toews, 2016). Hence, guidance is needed for how to develop learning activities that promote and integrate statistical and computational thinking practices. Statistical modelling tasks that introduce code-driven tools could provide an excellent learning situation for teaching integrated statistical and computational thinking (cf. Wickham, 2010). However, it is difficult to find examples from statistics education research literature that specifically explain how to *assess integrated* statistical and computational thinking. In our paper, we present a hypothesised framework that identifies eight potential observable integrated statistical and computational thinking practices.

ASSESSING THINKING PRACTICES FOR DATA SCIENCE

For an integration of statistical and computational thinking to be taught and assessed, the practices associated with the integrated thinking process need to be described in observable terms. Such descriptions could support teachers to design specific learning activities to promote so called *data scientific thinking* (Gould, 2021) and to use learners' responses and actions to evaluate and assess observed thinking practices. To develop a characterisation of the integration of statistical and computational thinking, we considered existing frameworks for statistical thinking (e.g., Wild and Pfannkuch, 1999), computational thinking (e.g., Brennan & Resnick, 2012), statistical computing (e.g., Woodard & Lee, 2021) and frameworks used with high school data science curricula. Based on these frameworks, we posited that it was more useful to interpret "integrated" as *connected* rather than combined. To illustrate our characterisation of integrated statistical and computational thinking, we created a framework based on the notion of connection as shuttling between contextual, statistical, and computational spheres (Figure 1).

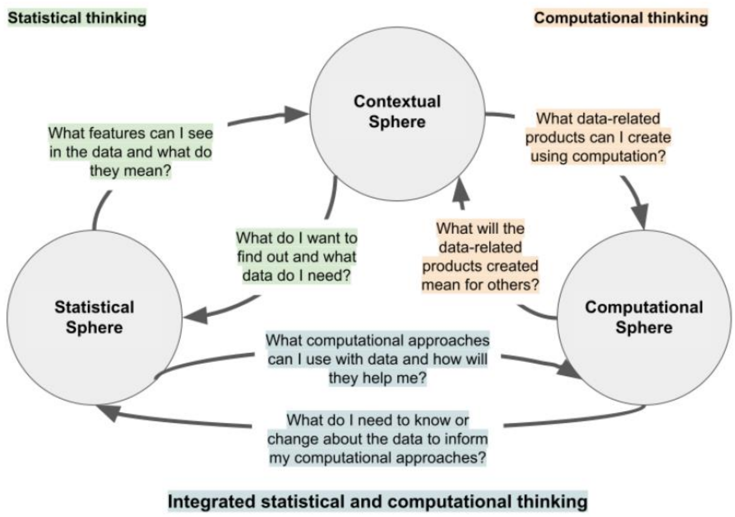

Figure 1: Characterisation of integrated statistical and computational thinking framework

Figure 1 draws on the work of Wild and Pfannkuch (1999). The characterisation framework theorises that to identify the thinking practices of learners, one should analyse sequences of actions in terms of how learners *move between* spheres, rather than describing thinking based on observing individual *computing actions* in learners' statistical work (cf. Woodard & Lee, 2021). The framework incorporates humanistic perspectives related to computational product design (e.g., Brennan & Resnick, 2012) to characterise shuttling between the contextual sphere and the computational sphere. Six questions are added as annotations for each of the six shuttles, with each question deliberately written from the perspective of a learner. We found that the complexity of describing the integration of statistical and computational thinking could be partially reduced by exploring two key questions for the development of a hypothesised framework to identify the thinking practices: (1) *In what ways do computational approaches support the development of statistical thinking?* (2) *In what ways does using data support the development of computational thinking?* (cf. Gelman & Nolan, 2017).

OUR HYPOTHESISED FRAMEWORK

Using a design-based research approach (e.g., Hoadley & Campos, 2022), four structured tasks were developed for teaching statistical modelling at the same time as introducing the programming language *R*. Each task focused on a different type of statistical modelling: classification modelling, randomisation tests, predictive modelling, and probability modelling. These four tasks were implemented with high school statistics teachers across four full-day face-to-face professional development workshops. Using Figure 1 and existing theoretical perspectives for statistical and computational thinking, we compared the teachers' observed thinking practices across the tasks to identify eight observable thinking integrated statistical and computational thinking practices (Table 1) that appeared to align with shuttling between the statistical and computational spheres. These thinking practices included what was observed in the teachers' actions and discussions, and what was not observed, that is, cases where a thinking practice could have been advantageous. We also considered our expectations and intentions as the task designers for the nature of thinking that could be supported or stimulated by different task features.

Observable thinking practice (OTP)	**Example from research tasks**
Connecting features of visualisations with statistical model components expressed computationally (OTP1)	Linking the band visualised around the fitted line to the line of code that determined the error component of the prediction model that generated prediction intervals (predictive modelling task)
Articulating a sequence of statistical modelling actions in computational terms (OTP2)	Describing the computational aspects of each modelling step using natural language, to match the computational actions captured in the screenshots from VIT online (randomisation test task)
Automating statistical modelling actions effectively using computational tools (OTP3)	Copying and adapting the given code for a probability simulation to explore the *Oh what a seed* problem situation and successfully producing a viable model for the card collecting promotion (probability modelling task)
Tinkering with statistical models productively to develop generalisations (OTP4)	Exploring new test statistics by changing aspects of the code for the model, leading to generalisations about the relationship between the features of the sample data, the test statistic, and the re-randomisation distribution (randomisation test task)
Restructuring or manipulating data for a statistical modelling purpose (OTP5)	Recognising the hierarchal nature of the data generated from the probability simulation and how it was represented in the computational tools (probability modelling task)
Developing algorithms by analysing data and selecting relevant variables (OTP6)	Developing a decision rule to classify grayscale photos as high contrast or low contrast by determining what features of the grayscale data to use for each photo and what "cut off" value to use (classification modelling task)
Obtaining data from digital sources and using data strategically to develop models (OTP7)	Sourcing data from the OMDb by using code to making queries to the API, then using the data from one query for training and the data from another query for testing (predictive modelling task)
Considering uncertainty and generalisability when using models as computational products (OTP8)	Discussing how random samples of pixels were being used to create the grayscale distributions for photos and the impact of this on model developed, specifically the decision rule (classification modelling task)

Table 1: Observable integrated statistical and computational thinking practices framework

Each observable thinking practice in Table 1 connects knowledge or use of data, models, and computational tools and is illustrated using one example from the research tasks. A general characteristic for each observable thinking practice is that the learner needs to draw on aspects of analysis, abstraction, and automation, the three core aspects of computational thinking according to Lee et al. (2011). Each observable thinking practice also demonstrates the need to *shuttle between spheres*, namely the *statistical* and *computational* spheres.

CONCLUDING REMARKS

The *observable integrated statistical and computational thinking practices framework* was developed from the teaching context of statistical modelling, and hence focuses on *connections* between the statistical and computational spheres based on *data* and *models*. To assess integrated statistical and computational thinking, teachers need to know what kinds of thinking practices to observe. Our framework may provide useful support to data science teachers with respect to creating statistical modelling tasks that introduce code-driven tools. Teachers could consider in what ways task design components encourage different observable thinking practices and future research will explore the usefulness of the framework in this regard (cf. Sentance et al., 2019).

References

Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking. In *Proceedings of the 2012 annual meeting of the American Educational Research Association (April 2012), Vol. 1, Vancouver, Canada*. AERA.

De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant, L., Cheng, L. Z., Francis, A., Gould, R., Kim, A. Y., Kretchmar, M., Lu, Q., Moskol, A., Nolan, D., Pelayo, R., Raleigh, S., Sethi, R. J., Sondjaja, M., … Ye, P. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application*, *4*, 15–30

Gelman, A., & Nolan, D. (2017). *Teaching statistics: A bag of tricks*. Oxford University Press.

Gould, R. (2021). Toward data-scientific thinking. *Teaching Statistics, 43*, S11–S22.

Hoadley, C., & Campos, F. C. (2022). Design-based research: What it is and why it matters to studying online learning. *Educational Psychologist*, *57*(3), 207–220.

Lee, I., Martin, F., Denner, J., Coulter, B., Allan, W., Erickson, J., Malyn-Smtih, J., & Werner, L. (2011). Computational thinking for youth in practice. *ACM Inroads*, *2*(1), 32–37.

Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician*, *64*(2), 97–107.

Sentance, S., Waite, J., & Kallia, M. (2019). Teaching computer programming with PRIMM: A sociocultural perspective. *Computer Science Education, 29*(2-3), 136–176.

Toews, C. (2017). Computational Inquiry in Introductory Statistics. *PRIMUS*, *27*(7), 707–724.

Wickham, H. (2010). Using visualisation to teaching data analysis and programming. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. International Statistical Institute.

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223–248.

Woodard, V., & Lee, H. (2021). How students use statistical computing in problem solving. *Journal of Statistics and Data Science Education*, *29*(sup1), S145–S156.