AN INFERENTIALISM-BASED FRAMEWORK FOR CAPTURING STATISTICAL CONCEPT FORMATION OVER TIME

ANNE PATEL
Waipapa Taumata Rau | The University of Auckland
a.patel@auckland.ac.nz

MAXINE PFANNKUCH
Waipapa Taumata Rau | The University of Auckland
m.pfannkuch@auckland.ac.nz

ABSTRACT

Statistics education researchers have been challenged to consider the theory of inferentialism in understanding concept formation in students. A critique of inferentialism is that no comprehensive method has been formulated to use the theory in practice. In this paper an inferentialism-based framework is presented that appears to be capable of explicating the development of statistical concepts during learning. By following six 11-year-olds' learning over several statistical modelling cycles using TinkerPlots, the framework was used to capture their interrogative cycles of noticing and wondering, giving and asking for reasons, and sanctioning and censuring, as well as oscillations between concretising language about actions and conceptualising language towards concept formation. Five teaching episodes occurring near the beginning of a 12-week learning sequence are used to illustrate how the framework might be able to capture student concept formation over time.

Keywords: *Inferentialism; statistical modelling; middle school students; theoretical framework; statistical concept formation*

1. INTRODUCTION

Many learning theories have been used in mathematics education research literature to explain the perspective that underpins a research study or how students learn. A continual criticism of statistics education research articles and research is the lack of attention to theoretical frameworks, including learning theories (Kaplan & Peters, 2023). Indeed, in an analysis of the theories used in statistics education research, Nilsson et al. (2018) recommended that researchers should be "explicit about their background theories and orientating frameworks" (p. 374) and that more attention to theoretical underpinnings would improve the quality and development of statistics education research as a scientific discipline. Some researchers (e.g., A. Zieffler, public comment at the Consortium for the Advancement of Undergraduate Statistics Education Research Satellite Conference, June 1, 2023), however, challenge the notion that the worthiness of research in statistics education should be judged and constrained by an insistence on couching the research within theoretical frameworks, because many interesting, far-reaching, and impactful findings can be elucidated without referencing theoretical underpinnings. Nevertheless, Bakker and Derry (2011) and Nilsson et al. (2018) argued that as statistics education research matured, there was a need for researchers to address the theoretical underpinnings upon which the research was based and suggested that researchers should investigate inferentialism (Brandom, 2000) as a viable theory for statistics education. In statistics education, Bakker et al. (2017) tested how the theory might be used by analysing one vocational student solving a problem, whereas Nilsson (2020) used inferentialism in a small-scale teaching experiment as a background theory, with his analysis focusing on the enactment of the *game of giving and asking for reasons*. As part of the debate about the potential of inferentialism as a theory for statistics education, this paper goes beyond the work of these researchers to demonstrate the *application* of an inferentialism-based framework that we developed to explore six 11-year-olds' learning as they engaged in two modelling tasks. The purpose

Statistics Education Research Journal, 24(1). https://doi.org/10.52041/serj.v24i1.714

of the paper is to show how the framework might be capable of capturing and characterising student concept formation within a statistical modelling environment over time.

2. LITERATURE REVIEW

There are many learning theories in existence. This section focuses on two prominent theories, constructivism and socio-constructivism, by giving a very brief overview and then discusses arguments that the theory of inferentialism (Brandom, 2000) may provide a way forward in helping to understand novices' statistical concept formation.

2.1. LEARNING THEORIES UNDERPINNING EDUCATION RESEARCH

Conflicting and complementary philosophical views of how knowledge is acquired permeate mathematics education research. Baroody et al. (2004) placed these views on a continuum based on the nature of knowledge in relation to one's view of authority, with direct instruction to promote procedural skills at one end and a "laissez faire" problem-solving or process-orientated approach for learning mathematics at the other end. Constructivism, with its many branches, is currently the dominant educational paradigm, which arose in response to behaviourist theories of learning. Constructivists moved away from direct instruction, the rationale being to prepare citizens for an increasingly industrialised and technology-driven society. They are concerned with ontology, what things are, and epistemology, how we come to know things (Radford, 2017). According to von Glasersfeld (1995, p. 18), constructivism is based on two founding principles: (a) knowledge is not passively received but built up by the cognising subject; (b) the function of cognition is adaptive and serves the organisation of the experiential world, not the discovery of ontological reality.

Newer sociocultural learning theories, such as socio-constructivism, view the acquisition of knowledge as dependent on the communities and society in which learning takes place, like an apprenticeship where skills, language, and behaviour are modelled. These theories drew on a wider and deeper range of societal cues and contexts that illuminated how, why, and what was taught and learnt. Radford (2018) pointed out that unlike constructivist theories of learning where autonomy is a "prerequisite" for learning, newer sociocultural theories viewed intellectual autonomy as a capability that is acquired through justifying one's reasoning to others. Radford's view aligns with Vygotsky's concept of a zone of proximal development, where learning takes place alongside a more knowledgeable other and autonomy is viewed not as a prerequisite but as a *result* of knowledge acquisition (Bakker & Derry, 2011; Radford, 2018). Moreover, Radford (2018) stated that knowledge can be thought of as hypothetical and only comes into existence when it is both used and useful.

The 21st century signalled a renewed interest amongst mathematics educators in developing new theories of learning and pedagogy, which Radford (2017) attributed to: an awareness of the assumptions underpinning Eurocentric conceptions of mathematics and learning; more interest in the role of culture, history, and societal issues such as inclusion, multiculturism, and political concerns; and overcoming perceived limitations of constructivist theories. Some researchers argued for a semantic theory of learning as an alternative to constructivist theories, because "interpretations are private [and therefore] it is not possible to directly observe them," and point to "the *inferred* quality" of an individual's learning experiences (Noorloos et al., 2017, p. 19). Noorloos et al. (2017) suggested an *inferential* view of the process of learning should be applied to mathematics and statistics education research, as constructivist terms such as "interpretations," "ways of knowing," or "conceptual schemes" used to describe knowledge acquisition concerning embodied forms and expressions of knowledge do not in themselves reveal the *process* of construction. For a full discussion on inferentialism and the limitations of constructivist theories, see Bakker and Hußmann (2017).

Within this milieu of debate about new theories for a post-industrial age, researchers (e.g., Bakker & Derry, 2011; Nilsson et al., 2018) proposed that Brandom's (2000) theory of inferentialism should be considered. Inferentialism is a philosophical, semantic theory about human activities and meaning making, which Bakker and Hußmann (2017) argued provides a resource for viewing the development of concepts in new ways and "has the potential to highlight the importance of agency of learners who participate in a social practice of giving and asking for reasons" (p. 390). From the inferentialist perspective, what distinguishes humans from other species is awareness of reasons, ability to articulate them explicitly, and to be *responsive to reasons*. In this way, learners are initiated into the language and the conventions of how to reason (Derry, 2017). Learners are enculturated into the practice of statistics in a problem-driven or inquiry-led, holistic learning approach whereby learners begin to deal with the content of concepts as well as recognise the conditions for their application through giving and asking for reasons and through making judgements and visible commitments.

An example of the inferential view of the learning process is where a learner may say that "*x* is the mean of five numbers" and then say "*x* is the middle number," both of which are commitments. The learner has also committed to a particular connection between the mean and middle number and endorsed an inferential connection by applying the two concepts (words in sentences) in a specific way (Derry, 2017). The correctness of what the learner has said now depends on the reactions of the other students and the teacher—the community—and the learner's responsiveness to them (Noorloos et al., 2017). The community may ask the learner for reasons to support their claims or commitments, which could lead the community to disagree with the learner for saying two incompatible things or to agree if the right reasons are provided. By the learner making their thoughts public, the community is encouraged to respond, perhaps to agree, disagree, or to ask further questions to either clarify or provide alternative reasoning. The aim for the community and learner is to reach a shared meaning, where agreed-upon inferences become an accepted "truth" or *new* knowledge among them. For example, the resultant discourse may cover the connections among mean, middle number, and median, namely the criteria for their use, the conditions of their application, and what the learner is entitled to say following their commitments (Noorloos et al., 2017). By being responsive to reasons arising within the community about how to reason with such concepts, the learner may begin to grow their knowledge about mean, middle number, and median. The approach is holistic because a concept cannot be reasoned about in isolation; rather a web of concepts is drawn upon to inferentially connect a concept to other concepts (Bakker & Derry, 2011).

Currently there are limited explanations of how learning could be investigated empirically and the pedagogy that theorists envisaged for inferentialism (Radford, 2017). Derry (2017, p. 409), however, does provide a picture of a teacher with an inferentialist approach as one who would attend to "what learners say and unpacking their attempt to articulate what they understand [and] may involve showing learners what they committed themselves to and what is entailed by their commitments." According to Derry (2017) inferentialism has the potential to (a) offer a fine-grained account of learning, whereas Nilsson et al. (2018) argue that inferentialism could (b) give insights into concepts in use; (c) provide a holistic account of the process of concept formation; (d) elucidate the interplay between technology and human-based decisions; and (e) demonstrate the connecting of statistical and contextual knowledge. Therefore, as part of the debate about inferentialism, the search for potential post-industrial theories of learning and pedagogy, and the challenge to statistics education to consider inferentialism as a theoretical resource for epistemological reflection on student learning (Bakker & Derry, 2011; Nilsson et al., 2018), we offer our attempt at using inferentialist ideas to investigate the process of statistical concept formation during learning.

2.2. AN INFERENTIALISM-BASED FRAMEWORK

As part of a larger study that examined the *learning processes* involved in statistical modelling, an inferentialism-based (IB) framework (Figure 1) was developed to explain and interpret *what* statistical concepts students were encountering and navigating when modelling, and to illustrate *how* this occurred. To elucidate how students might encounter and navigate concepts during modelling, we constructed the IB framework that drew on three main sources: (a) the theory of inferentialism (Brandom, 2000); (b) the pedagogical practice of noticing and wondering (Shaughnessy, 1997); and (c) the idea that reasoning coordinates knowledge and actions (Heusdens et al., 2019). Note that when inferentialism theory uses the term inferences, it is not the same as drawing statistical inferences. The IB framework comprises interconnected inferential cycles involving multiple interrogative cycles of giving and asking for reasons, sanctioning and censuring, and noticing and wondering, and multiple oscillations between concretising and conceptualising, which are now elaborated upon.

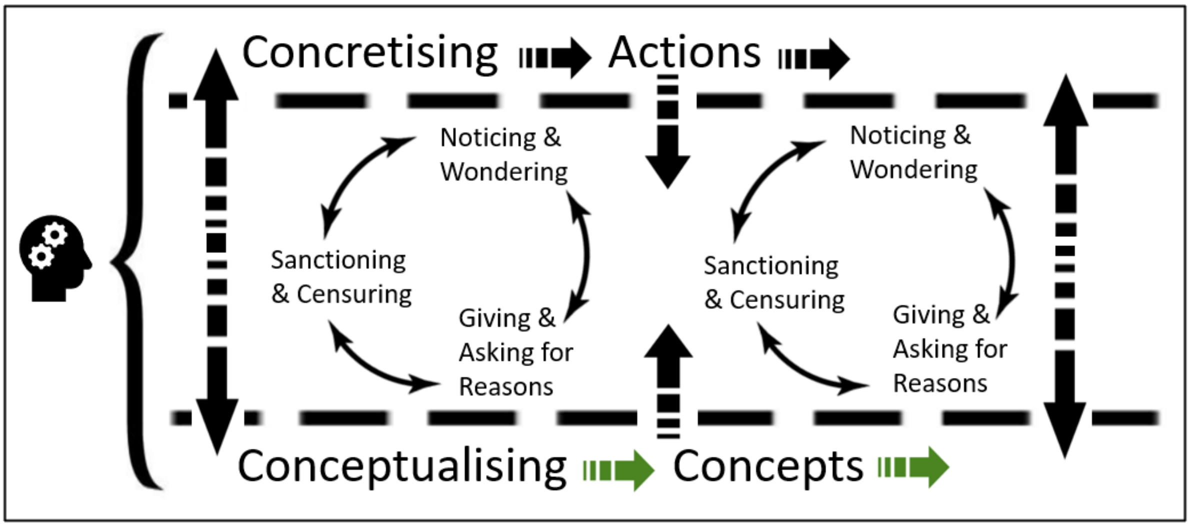


Figure 1. An inferentialism-based framework (Adapted from Patel, 2022, p. 75)

Giving and asking for reasons Inferentialists view cognitive activity as both social and individual. Within the social game of *giving and asking for reasons*, inferentialists understand learning takes place within individuals as they form and use a "web of reasons" that reflects both conceptual understanding and the mastering of a practice (Bakker & Derry, 2011; Brandom, 2000; Noorloos et al., 2017; Taylor et al., 2017). A central tenet of inferentialism is the role language plays in one's social ability *to give and ask for reasons*. Brandom (2000) viewed these exchanges as a human desire for truth seeking, which facilitates knowledge building where concepts are created and used in terms of their reasoned and inferred connections (Bakker et al., 2017; Bakker & Derry, 2011). Inferentialism elucidates the role of language in reasoning where the meaning of words is explicated through how they are used in social practices (Noorloos et al., 2017). Claims or commitments require supporting reasoning and "what counts as valid reasoning, adequate judgment, or correct application of concepts depends on the norms being used in a particular practice" (Bakker & Derry, 2011, p. 12).

For education in schools, the social norms of the classroom dictate what is inferred or questioned between individuals. To adopt an inferentialist approach, a teacher needs to engage the classroom participants in the linguistic activity "game of giving and asking for reasons" (Brandom, 2000, p.189). Noorloos et al. (2017) contended that inferentialism explains the formation of concepts "in terms of the inferences individuals make in the context of an intersubjective practice of acknowledging, attributing, and challenging one another's commitments" (p. 437).

Sanctioning and censuring In line with socio-constructive theories, inferentialism is based on social norms that dictate what is inferred or questioned between individuals in a "game." Brandom's (2000) theory of how the "game" is played has speakers keeping "score" of the commitments that each is allowed, or not allowed, to make according to social norms, where the viability of a claim or commitment depends on the response of the participants in the dialogue and on the person making the commitment (Noorloos et al., 2017). In a public conversation interlocutors can engage with and interrogate commitments, encourage responses, clarify reasoning, and either agree (sanction) or disagree (censure). In this way, a collective meaning or understanding of agreed-upon inferences become an accepted "truth." Lines of reasoning not agreed upon or censured between speakers are likely to deteriorate over time. Thus, inferentialism places an "emphasis on reasoning and norms and its conceptualization of the social aspects of these phenomena" (Noorloos et al., 2017, p. 451).

Noticing and wondering Inferentialism views the ability of individuals to give and ask for reasons of others as facilitating the acquisition and utilisation of knowledge *mediated* by the world, including technology and society around them. In other words, one notices, wonders, and attempts to make sense of what one perceives about the world. In pedagogical practice, *noticing and wondering* (Shaughnessy, 1997) is a pragmatic method of drawing students' attention to the statistics world they are navigating by prompting them to make inferences through "noticing characteristics of displays" and warranting or justifying their "inferences about questions or claims related to their inquiry" (Lehrer & English, 2018, p. 248). For example, students may *notice* the features of chance distributions, randomness in simulations, or increased stability under large trials, which prompt them to question or infer about variation in distributions based on what they already know. Students viewing the outputs of their models in simulations naturally *wonder* about what they *notice* and may publicly offer reasons to explain what they see and their thinking.

Concretising and conceptualising Brandom (2000) highlighted that concepts and actions drive learning. Indeed, Heusdens et al. (2019), in their research in a vocational setting, fused an action-based theory of learning (Van Oers, 1998) with inferentialism to demonstrate a continuous oscillation between actions and conceptual development. They theorised two sensitising language constructs from their data: (a) *conceptualising*, which is characterised as articulations about general concepts that promote a movement towards understanding; and (b) *concretising*, which is characterised as articulations interpreting the situation about what is happening or what should be done that encouraged a movement towards action. Combined, these language constructs gave meaning to the situation or task and were used as evidence that concepts and actions were tied together during learning (Heusdens et al., 2019). However, their research used domain-based questioning of trainee chefs' knowledge, elicited by an experienced interviewer, which took place *after* rather than during learning. In a similar vein, Bakker et al. (2017) argued that when people are coordinating knowledge and action, the glue that holds them together are reasons and inferential relationships. Their study on one vocational student solving a problem, as part of his internship in a hospital laboratory, showed how the coordination of knowledge and actions involved *webbing*; for example, where connections are made between statistical object (e.g., parameter) and class of problem (e.g., experiment), and meaning is reshaped and recreated in action.

Further notes on the IB framework Figure 1 depicts the interrogative space (within the two horizontal dashed lines) in which giving and asking for reasons, sanctioning and censuring, and noticing and wondering continually operate. Any part of the interrogative cycle *may* result in an action (upward dashed arrow into the concretising actions region) such as changing a representation, setting up a factor for a model, or running a simulation. That is, an action is the learner *doing* something. Once the action is undertaken, the learner moves back into the interrogative space (downward arrow). Similarly, any part of the interrogative cycle may activate the process of concept formation (downward dashed arrow into the conceptualising concepts region) followed by a return to the interrogative space (upward arrow). In the conceptualising concepts region the horizontal arrows are coloured green to indicate that concepts are in the process of formation and that they will deepen and grow over time, yet they are continually in formation during the activation of the interrogative cycles. In contrast the concretising actions region is an accumulation of many actions running alongside the continual operation of the interrogative cycles. Thus, when the interrogative space is activated there is a constant shuttling between the concretising actions region and conceptualising concepts region.

In the development of the IB framework, student data was used to inform and trial ideas about how to practically capture students' interaction with the technology and other people and to capture the formation process of new concepts. In particular, the pedagogical practice of noticing and wondering about statistical representations was not mentioned in the inferentialism literature, yet the data clearly pointed to representations being an important stimulus to interactions and to the concretising and conceptualising processes. Thus, the IB framework evolved from knowledge of the inferentialism literature and from the back-and-forth dialogue and actions of novice students between themselves and with the researcher as they engaged in building chance-based models to investigate situations using different modelling contexts.

In education there is an assumption that knowledge involves representations or re-presentations in the wider sense, and that to learn is to construct representations, whereas inferentialism states making judgements is the fundamental unit of knowledge rather than representations (Bakker & Derry, 2011). For example, learning statistical concepts occurs through reasoning that makes representations meaningful. Thus, the interrogative cycle highlights a central tenet of inferentialism that concepts are formed in relation to other concepts, not in isolation, within the social practice of giving and asking for reasons and making judgements. Although inferentialism has begun to be used to examine how concepts and knowledge are acquired and used (e.g., Nilsson, 2020), there still remains a gap in the inferentialism literature about how student reasoning and actions facilitate the acquisition and use of statistical concepts *during* learning. Currently there is no comprehensive applied framework based on inferentialism that can examine students' conceptual development through analysis of their language and actions. Therefore, the purpose of this paper is to show how the IB framework might be able to capture the formation process of new concepts over time.

3. BACKGROUND TO RESEARCH AND PARTICIPANTS

This paper is set within a wider two-year study examining six novice 11-year-old students' and a researcher's actions and discourse (Patel, 2022). The overall learning objective was to introduce students to statistical modelling and to enhance and extend their probabilistic thinking (Patel & Pfannkuch, 2018). The class teacher selected the six student participants from the 29 students who consented to be involved in the study based on their ability to explain their thinking, good verbal skills and varied mathematical skills. A learning sequence to model phenomena was developed using both the Data Analysis and Sampler tools in *TinkerPlots* (Konold & Miller, 2011). The students had no prior experience with *TinkerPlots* or distributional ideas. They were considered to be high ability in their mid-socio-economic school. The students worked in pairs: Krista and Mary, Ali and Leo, and Nico and Dan (pseudonyms). Student data (audio, screen captures and artefacts) were micro-analysed to capture their interactions between each other and with the technology during a series of modelling tasks over 12 two-hour lessons, where each lesson occurred once a week. Note all student visualisations in *TinkerPlots* have been re-created for the purpose of clarity in this paper.

3.1. ANALYSIS AND CODING

The analysis and coding were carried out by the researcher, the first author. The analysis and coding were conducted on both the students' and researcher's utterances because the researcher was a participant in the learning process, for example, prompting students to notice and wonder, asking for reasons, and, where appropriate, either sanctioning or censuring (cf. Derry, 2017). By coding the researcher, her part in the learning process is recognised as well as her potential influence on guiding students towards actions and concept formation.

At a micro-analysis level, the verbatim, transcribed data was used alongside screenshots of what the students were looking at, using, or building in *TinkerPlots*. Beginning with the students' verbalisations and actions, the analysis of the data focused on the students' perspective. With progressive iterations through the data line by line, the coding was enacted "through the constant comparative method" (Charmaz, 2014, p. 181), where every part of the data is constantly compared with other parts of the data. Some parts of the data, where the students encountered and resolved issues or demonstrated interesting reasoning patterns, were subjected to intensive analysis to understand more fully students' emergent concepts. This involved unpacking probable concept formation via inferred links or connections based on the students' language and related actions. The researcher's interpretations of the student data were debated, challenged, and discussed with the second author.

The coding began by searching for questions, indicated by question marks and associated with key words, such as what, when, why, and how in the verbatim transcripts. These were coded as "asking for reasons" or "wondering." Differentiating between these required considering the intention and *history* of the speakers and implicitly interpreting their intention based on the context of the discussion. The responses could be motivated by the problem, simulated or real data, the model, or the context. The response was coded as "wondering" if the speaker responded by being unsure through the use of words such as maybe, using higher intonation indicating doubt, or asking another person to wonder about a situation. The response, however, was coded as "giving a reason" if it contained a statistical or contextual response that, based on past utterances, indicated "inferential uptake" and not something that "a parrot could be trained to perform" (Nilsson, 2020, p. 54). Further key words such as "because," "'cos," "if," and "when," associated with providing a reason were searched for and coded as "giving a reason" if they were related to the task, technology, statistics, or context. Usually, after the students performed an action in *TinkerPlots*, instances of noticing were coded, which could be questions or statements such as "look" or "wow" that may have focused others' attention on what had been noticed.

Sanctioning, censuring, and wondering were also observed in the following three scenarios and consequently classified: (a) as sanctioning when approving other speakers' utterances; (b) as censuring, repeated as a query, which involved consideration of intonation, with higher intonation indicating doubt or a question; and (c) as wondering or thinking about the issue through repeating a phrase or word and sharing their concerns publicly with others. Sanctioning utterances, in particular, were usually accompanied by positive language; for example, "That's the thinking" or "You did it!" These also included subtle sanctioning, such as repeating or re-voicing others' reasoning, which generally involved more precise or meaningful language, particularly by the researcher. Such interactions gradually inducted students into the language and reasoning that would be sanctioned. As Noorloos et al., (2017, p.448) stated:

> In interactions, people test their inferences and their commitments and in this way are able to alter their understanding of the words they use. Because the status of their utterances develops in the context of how they are assessed by others, it is not necessary that an explicit thought process underlies each piece of activity; the acts are rational if they make sense to others in connection with the other commitments of the agent.

Therefore, if a student *only* did an action in response to another student's utterance, the action was coded, for example, as sanctioning because, based on previous exchanges between the two students, it made sense. Harsh censuring utterances were also identified; however, subtle censuring proved more difficult to code. For example, utterances such as "But what about..." were dual-coded as both censuring and giving a reason if it was followed by a related inference. These subtler forms of censuring had to be considered in relation to data over time, as well as various apparent power relationships between discussants. Indeed, these utterances were usually associated with a refocusing or redirection about what to do and illuminated *how* students proceeded through a model cycle.

Following the coding of the transcripts for the interrogative cycle, periods of continuous, intense reasoning exchanges were examined when statistical concepts or actions were discussed. The analysis shifted focus to the *language* the students used to infer about concepts (i.e., conceptualising), and to identify what they noticed, reasoned about, and experimented with during these exchanges. Brandom (2000) writes "expressing something is Conceptualising it: putting it into conceptual form" (p. 27) or *navigating* the concept. Given the age of the students, their exposure to statistical language could be characterised as experimental. Therefore, they were *experimenting* with language—for example, the words random and sample—that they had heard used in their daily lives. Indeed, at this age, their use of these words may not coincide with the words' statistical meanings. Likely concepts, which in many cases were just inklings of concepts (e.g., model fit) or their informal precursors, were recorded alongside the transcripts.

Because concepts are not conceived in isolation but rather as inferred relationships to other concepts, the identification and tracking of the formation of new concepts needed to be considered holistically. In particular, the formation of the concepts of model fit, causes of variation, and sample space were tracked in the analysis, yet each one draws on a network of concepts integral to reasoning with and about them. Model fit, for example, is an expectation about features of the data or distribution one expects to see (Konold & Kazak, 2008). Students evaluating their simulated distributions against the real data distribution draw on similarities in shape, central tendency, and spread, all of which might not be articulated verbally. However, over time, the formation of the concept of model fit can be tracked through what the students, with oversight and participation from the researcher, commit to, challenge, reject, or collectively accept as "true." Similarly, for causes of variation, the factors that may contribute to the variation in the data distribution can be tracked through what students consider to be or not to be causes based on their articulation of their contextual knowledge or experiences and what they commit to, challenge, reject, or accept as potential causes. The concept of sample space is considered fundamental in promoting growth in probabilistic thinking, yet is very difficult to grasp (Chernoff & Zazkis, 2011). Within a modelling environment, students' conjectures about the sample space are realised in the random generators (e.g., spinners) that are built as part of the model and which are experienced as an inferred relationship among the model, simulated distributional patterns, and real data distribution. Thus, the identification and tracking of the formation of sample space conceptions is based on what students use, accept, challenge, and reject as sample spaces in their building of the models and their consequent collective acceptance or rejection of the simulated distribution sample space.

Finally, examples of concretising language were coded, which Heusdens et al. (2019) stressed are not actions, but *language use* leading to possible actions. The students' actions in *TinkerPlots,* observed in the videography, were recorded alongside their discourse, such as the creation of a plot or locating the mean of a sample.

4. ILLUSTRATIONS OF THE USE OF THE IB FRAMEWORK

To illustrate the application of the IB framework, two statistical modelling tasks, the Moon Hopper Task, and the Cookie Cutter Task, are used. To understand the interactions between the participants and the formation process of new concepts, the background to the tasks is described followed by learning episodes that start with an overall view in Episode One. From Episodes Two to Four, different aspects of the IB framework are introduced and built on to enable the reader to fully comprehend the coding and use of the IB framework, culminating in a full description in Episode Five.

4.1. THE MOON HOPPER TASK

Background to Moon Hopper Task The Moon Hopper Task, involving repeated measures of an object, was based on tasks developed by Konold and Kazak (2008) and Lehrer (2015). The first part of the task occurred in Lesson 1 with the whole class (*n* = 29) from which the six students were selected. It required students to measure two attributes of a moon hopper—an inflatable ball designed to sit on and bounce (cf. Konold & Kazak, 2008). The students individually measured the circumference of the moon hopper and the diameter of the handle with instruments of different precision—a measuring tape and electronic callipers. They were instructed to write their measurements on Post-it notes. Two students were asked to draw a scale and arrange the class Post-it notes to create a plot for the circumference of the moon hopper on the whiteboard (Figure 2). The researcher led a discussion about the resultant distribution and then discussed possible causes for the variation in the data.

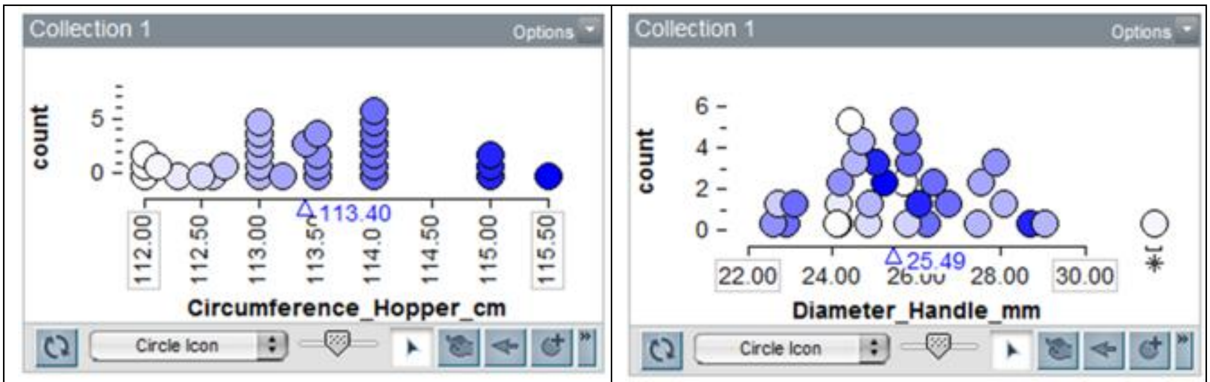
Figure 2. Arranging repeated measures of the moon hopper circumference

After the whole class was introduced to the capabilities of the *TinkerPlots* analysis tools in Lessons 2 and 3, the second part of the Moon Hopper Task occurred in Lesson 4 with the six selected students who worked in pairs. The students were asked to create plots in *TinkerPlots* of the moon hopper circumference (in centimetres with one decimal place) and the handle diameter (in millimetres with two or three decimal places). After the students had constructed their plots in *TinkerPlots*, the researcher led a group discussion about the differences and similarities seen in the two data sets, recalling the possible causes for the variation in the data that the class had discussed in Lesson 1.

We now illustrate how the IB Framework may capture the beginnings of statistical concept formation. The first episode gives an overall sense of the IB framework by presenting a dialogue between the researcher and the whole class, whereas the second episode illustrates part of the interrogative cycle, giving and asking for reasons and sanctioning, and the beginning of concept formation about causes of variation, by presenting a dialogue between the six students and the researcher.

First episode The first episode occurred in Lesson 1, the first week of the study, with the whole class when the researcher discussed the distribution of the moon hopper circumference measurements, which were on the whiteboard (Figure 2). The students noticed that one of the circumference measurements was extremely different from the others (Noticing). The researcher asked, “What’s going on here?” (Asking for a reason). The students suggested that the measurement was an error or “mistake” (Giving a reason, Conceptualising) because the circumference of the moon hopper could not be 80cm, as they could see all the other measures were between 112cm and 115cm (Giving a reason). The researcher used the language “dirty data” to describe the measurement error and asked, “Should we try and fix it, or discard it?” (Concretising language). Students in the class responded that it should be discarded, and the researcher removed the Post-it note from the board and tore it up (Action). This episode illustrates the inferential connections students may have made between concepts, such as measurement error, signal in data (central tendency), distance from signal, and error judgements. By using language and related actions and judgements for cleaning data, the “new” concept of “cleaning data” may be strengthened. The next inferential cycle (Figure 1) was continued when the researcher asked the students, “What do you think the real circumference of the moon hopper is?” (Wondering). This question focused students’ attention on features of the shape or centre of the distribution of measures shown by the Post-it notes (Noticing). The students agreed that the centre of the data, approximately 113.5cm, was likely to be the “real” circumference according to the measurement data (Giving a reason, Conceptualising). Their reasoning was sanctioned by the researcher, and in this way an interrogative process between concretising (language toward actions) and conceptualising (language about concepts) is set in motion.

Second episode In Lesson 4 (fourth week of the study), in preparation for building their first chance-based model in *TinkerPlots*, the six students in the study created plots of the moon hopper measures (Figure 3) and then were shown how to apply the average tools to represent the centre of the data. The mean is depicted by a blue triangle.



(a) Circumference of moon hopper (80cm measurement removed)	(b) Diameter of moon hopper handle

Figure 3. Plots of moon hopper measurements in TinkerPlots

Referring to both data plots, the researcher began a discussion with the intention of eliciting possible causes for the variation in the distributions of measurements, similar to the repeated measures task in Konold and Kazak (2008). In line with Pratt’s (2011) contention that students’ intuitive causal reasoning should be harnessed in statistical modelling, the researcher prompted the students to examine the distributions and think of causes that might underpin the variation observed. Thus, the interrogative cycle starts with the student *action* of producing two graphs (Figure 3) and Dan *noticing* the differences in distributions in Figure 3, which the researcher builds on (Table 1, Row 1). The backward and forward dialogue in the interrogative cycle then starts feeding into the process of the formation of the concept, causes of variation, within different students, which is indicated by X in Table 1.

Table 1. Interrogative cycle coding

Row	Interrogative cycle	Who	Utterance	Causes of variation
1	Asking for a reason	R:	Can you suggest a reason why the handle (data) looks more like a pyramid, or a triangle and the Hopper (circumference) has gaps?	
2	Giving a reason	Dan:	Cos, some people with the handle, sometimes they might have squeezed it, so you would get different results for that, and with the Hopper, you could have it (measuring tape) lower or higher on the Hopper. So, it could be out of place.	X
3	Sanctioning	Nico:	Because it (tape) could be out of place.	
4	Giving a reason	Nico:	Because the Hopper is bigger, bigger data variation.	X
5	Sanctioning Asking for a reason	R:	What do you think is causing the bigger variation?	
6	Giving a reason	Ali:	The size of it, so there is more to measure.	X
7	Sanctioning	Leo:	Yeah.	
8	Giving a reason	Leo:	More room for error than that little device thing (electronic callipers) we had, you just stuck it on. Lots of different variables that could happen.	X
9	Giving a reason	Ali:	If they measure the moon hopper, it (measuring tape) might have been down here or up there.	X
10	Sanctioning	R:	Right, so the angle might have been different.	
11	Giving a reason	Ali:	Probably how tight they hold the tape around the hopper.	X
12	Giving a reason	Nico:	If you squeeze it, it might be crooked.	X
13	Sanctioning	Dan:	Yeah, angle means position.	X

Overall, as illustrated in this episode, the four students are classified as beginning the process of forming a new concept, *causes of variation*. Note that Nico in Row 3 *sanctions* Dan's *reason* in Row 2 by revoicing his last sentence, and then he goes on to offer a further *reason* in Row 4. The researcher response in Row 5 is double-coded because she simultaneously revoices or *sanctions* Nico's response in Row 4 and *asks for a reason*. In Row 13 Dan echoes and *sanctions* the researcher's naming of a potential variable, angle (Row 10), and *sanctions* Nico's *reason* in Row 12.

The researcher then clarified that position and angle could be thought of as two different causes; position was vertical displacement from circumference, whereas angle referred to tilt of measuring tape. She wrote on the board the possible causes of variation suggested by the students: position, angle, pressure, starting point. This action served to sanction and further assist students' conceptualising about causes of variation. To prompt students to think about the numerical size of the factors causing the variation, the researcher then steered the discussion to possible values or sizes for each of the causal factors. Under guidance, the students then built their first chance-based model in *TinkerPlots* (e.g., Figure 4), where the first spinner yielded the average measure, and the other spinners yielded contributions to the variation in the measurements from students' identified causes. In other words, their ideas became physical tools in an additive model that could be modified, trialled, tested, and critiqued as they ran simulations and compared the resultant distributions to the real data distribution (e.g., Figure 5).

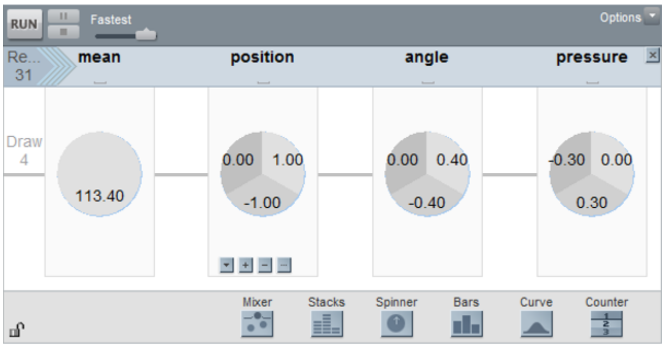
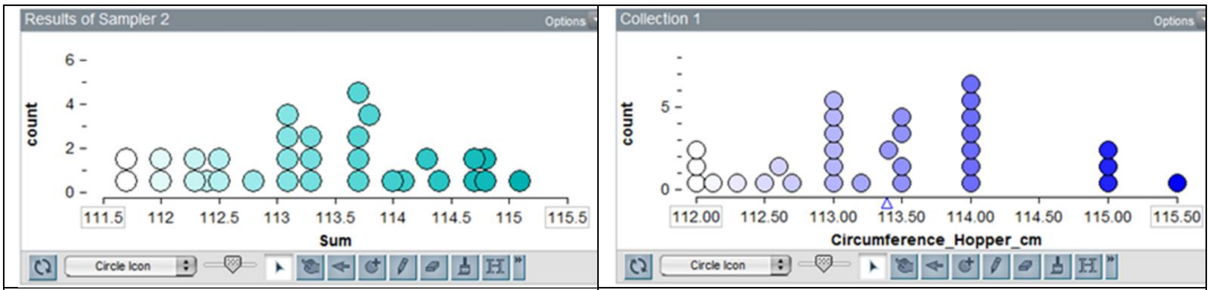


Figure 4. Krista and Mary's final model of repeated measures of the moon hopper circumference



(a) Simulation | (b) Moon hopper circumference distribution

Figure 5. One simulation for Krista and Mary's model (a) and the real data distribution (b)

By asking students to compare and reason about the differences in the two distributions, the researcher raised awareness about causes of variation. The students intuitively drew on their prior experiences or actions measuring the moon hopper and offered valid contextualised reasons for the *causes of variation* in the data. The reasons given were sanctioned in order to arrive at a shared understanding about the main possible causes of variation. The reasoning was sanctioned, either by agreement or repetition or expanding on the reasoning towards shared definitions and more robust inferences. The students used the measurement context and physical characteristics of the measurement tool and the moon hopper to *explain the variation in the distribution*, a key stage in the statistical modelling process.

4.2. THE COOKIE CUTTER TASK

Background to Cookie Cutter Task The Cookie Cutter Task, Lesson 5 (fifth week of the study), was designed as a model eliciting activity (Lesh & Doerr, 2003) and was based on the work of Konold and Harradine (2014) for process measurement data. The purpose of the Cookie Cutter Task was to determine which Kiwi-shaped cookie cutter, big or small, produced more consistent cookies to sell at a gift shop to raise money for a school trip. The problem was to investigate the weights of cookies produced using two different mechanisms, a 10-cm and 14-cm cookie cutter. The first part of the task required the students to: (a) manufacture cookies using playdough and record their weights; (b) analyse two sets of real data in *TinkerPlots*; (c) build and test two cookie weight models; and (d) analyse the simulated data and judge model fit. Essentially, they were comparing two groups of data: firstly, the two real data sets, and secondly, the real data with the corresponding simulated data for model fit. It was anticipated the models built would be similar in structure (i.e., an error structure) to their previous moon hopper model.

The students were informed that they had to attempt the task by themselves with no help from the researcher unless they were really stuck. They were not told how many of each size cookie to make. The six students worked as a group to manufacture and weigh around 20 cookies of each size (Figure 6).

Figure 6. Manufacturing large and small Kiwi-shaped cookies

The students then created side-by-side dot plots of the cookie weight data (Figure 7). They used the average tool, selecting the mean, to represent the centre of the groups, and put the numerical value of the mean on their plot (blue triangle and text in Figure 7).

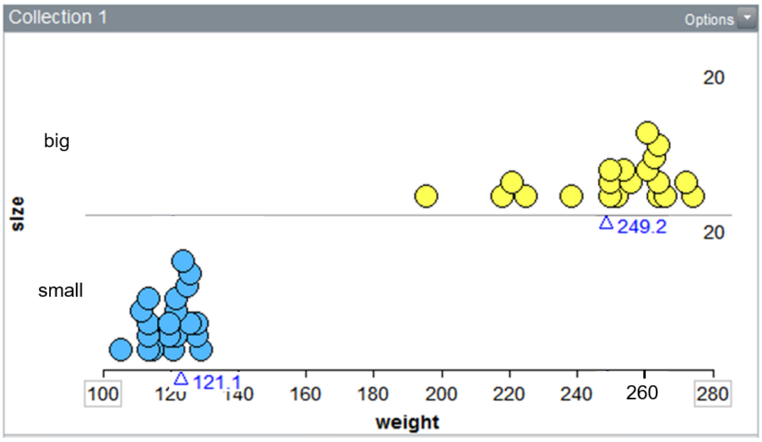


Figure 7. Plots of cookie weight data, by size

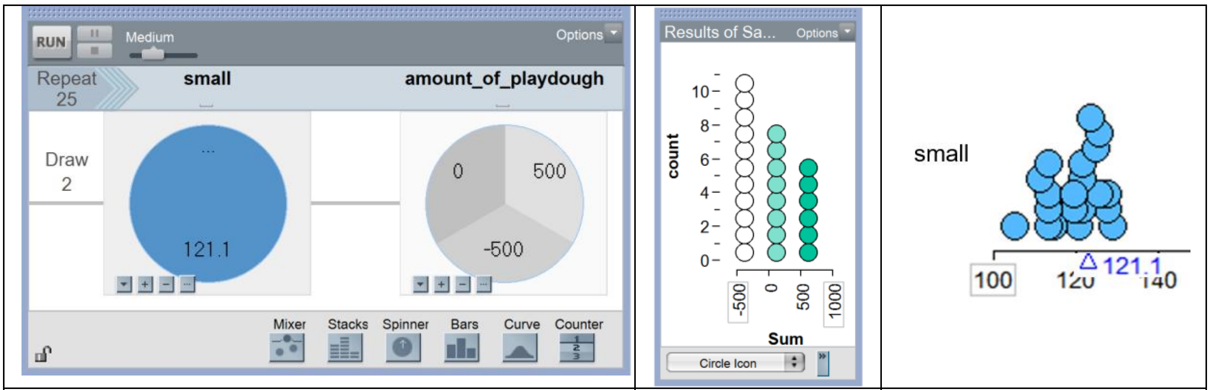
For the second task we have chosen to focus on Ali and Leo because they showed a trial-and-error approach for estimating the amount of variation each factor contributed and for assessing model fit. The following three episodes cover parts (c) and (d) of the task for the small cookie weights only. Over the episodes, we gradually build towards how the IB Framework may capture the beginnings of statistical concept formation for Ali and Leo.

Third episode In this episode (Table 2), Leo thinks of a factor based on his experience of making the cookies, that is, the data context. They create a spinner in a back-and-forth dialogue by using *actions* in the *TinkerPlots* Sampler, run a simulation, and then notice, wonder, give reasons for, ask for reasons about, sanction, and censure the result of their actions. In their dialogue they start the process of inferential concept formation.

Table 2. Interrogative cycle, and action and concept oscillation coding

Row	Action	Interrogative cycle	Who	Utterance	Concept(s) in process of formation
14	Creates spinner	Wondering	Ali:	Hurry up, think of a variable. Tell me it.	
15		Wondering	Leo:	Aah 150?	
16		Censuring	Ali:	No.	
17		Giving a reason	Ali:	Think of something that could affect the data.	Causes of variation
18		Wondering	Leo:	Oh, maybe how much playdough we put in? Playdough amount.	Causes of variation
19	Enters title for spinner.	Sanctioning	Ali	Amount of playdough. (Spells title of spinner out loud.)	
20		Sanctioning	Leo:	Ha-ha, great spelling.	
21		Asking for a reason	Leo:	Why did you put underscore?	
22		Giving a reason	Ali:	I didn't, when you do space, it comes up (automatically).	
23		Wondering	Ali:	And by how much do you think that (amount of playdough) would affect the data?	Sample space
24		Wondering	Leo:	Aah by 500%?	
25		Sanctioning	Ali:	OK.	
26		Censuring himself	Leo:	No, I reckon no. I reckon it will affect it by …	Sample space
27	Creates spinner with 3 sections, enters 0, 500, -500, (Figure 8a), presses run.	Censuring Wondering	Ali:	Wait, just run this, and see how it goes.	
28		Sanctioning	Leo:	OK.	
29		Noticing	Leo:	(Looks at outcome, Figure 8b, and compares to actual data in Figure 8c) Ooh, hmm…	Model fit
30		Wondering	Leo:	I would say, let's say oh, I would say…	

In episode three, Ali takes on the role of the teacher and thus takes the *action* of creating a spinner (Row 14). He asks Leo for a variable (Row 14), which we interpret as *wondering* what Leo's response will be. Leo's response in Row 15 has high intonation, indicating doubt and uncertainty. Therefore, Leo's response is classified as *wondering*, which Ali immediately censures (Row 16) and gives a reason for that censure (Row 17). After Leo suggests amount of playdough as a variable but still with doubt (Row 18), the *action* of Ali entering Leo's suggestion as the title of the spinner is an implicit *sanctioning* of his suggestion (Row 19). Similarly, Ali *sanctions* Leo's suggestion of 500% (Row 25) but subtly *censures* by his *action* of entering 0, -500, +500 on the spinner (Row 27). Leo also *censures* himself (Row 26).



(a) Model	(b) Simulation	(c) Small cookie weight distribution

Figure 8. Ali and Leo's small cookie weight model with one factor (a), resultant simulation (b), and real data distribution (c)

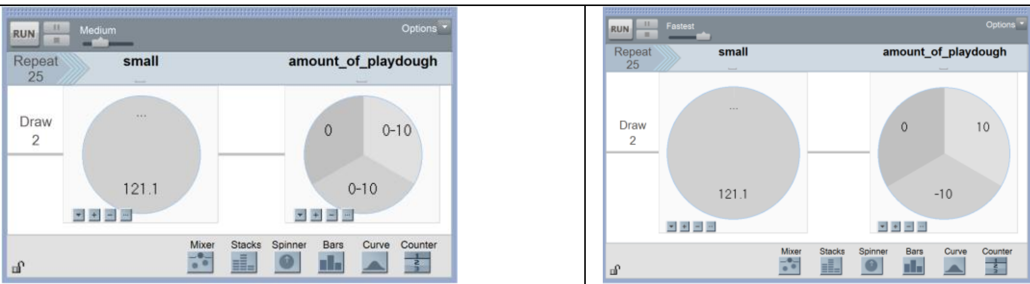
After running the simulation, both Ali and Leo (Row 29) realised that their estimate of 500 for the size of the deviation from the mean for the factor, "amount of playdough," was incorrect. Running a simulation gave them quick feedback, enabling them to rethink the magnitude of the size of the variation that might be contributed by the factor or variable. Concepts in the process of formation are *causes of variation*, *sample space,* and *model fit*, which are interconnected and enacted in an oscillation between actions and conceptualising through the interrogative cycles.

Fourth episode In this episode (Table 3), Ali and Leo's realisation about the incorrect magnitude of the size of the outcomes assigned to each spinner section not only leads them to reconsider the size but also to reconsider how a range of outcomes could be actioned in a spinner, not the three sizes for each spinner (+x, 0, -x) as all the students had done for the moon hopper model (Figure 4). Their back-and-forth dialogue facilitated them to increase the number of divisions on the spinner, which seemed to progress them towards a more sophisticated conception of continuous measurement to represent variation. The yellow shading of utterances highlights the notion of what is meant by conceptualising language that may promote a movement towards the process of the formation of a concept, whereas the blue shading refers to concretising language that interprets the situation that may or may not lead to action.

Row 33 is classified as *wondering* because there is doubt, but in Row 35 Leo backs up his suggestion by *giving a reason*, which Ali sanctions (Row 36). Leo's utterances in Rows 33 and 35 are examples of *conceptualising language* that may promote concept formation: *sample space*, *variation* (idea of range), *variation from mean*, and *distribution.* Leo *wonders* whether it is possible to put 2 into the spinner divisions (Row 39), which the researcher *censures* and then uses *concretising language* about what *action* to take (Row 40). Ali *sanctions* the researcher's suggestion with *concretising language that leads to an action* (Row 41). Leo *notices* the result of the action (Row 42) and *wonders* about entering other numbers using *conceptualising language*, suggesting an expanded notion of *sample space* (Row 43) mediated by the *TinkerPlots* environment.

Table 3. Interrogative cycle, oscillation, and conceptualising and concretising language coding

Row	Action	Interrogative cycle	Who	Utterance	Concept(s) in process of formation
31		Censuring	Ali:	Yeah, you said 500%	
32		Giving a reason	Leo:	I couldn't think of what else to say.	
33		Wondering	Leo:	So, um, I reckon it will affect it by um, by about, between one and 10 grams.	Sample space, Variation
34	Enters 0-10 on spinner (Figure 9a)	Sanctioning	Ali:	OK.	
35		Giving a reason	Leo:	Yeah, because if I say five (grams), it (the variation that I would see in the resultant plot) would be different.	Variation from mean, Distribution
36		Sanctioning	Ali:	Perfect.	
37		Asking for a reason	Leo:	Why did you put zero to 10?	
38		Giving a reason	Ali:	To turn it into a circle.	
39		Wondering	Leo:	You can't put 2 in there can you?	
40		Censuring	R:	No, split it into separate divisions.	
41	Changes numbers (Figure 9b)	Sanctioning	Ali:	Ok, you got to do -10 and 10.	
42		Noticing	Leo:	But what about the 8s, 9s, 7s?	
43		Wondering	Leo:	We'd have to go into them all wouldn't we?	Sample space



(a) First attempt at range of values | (b) Second attempt includes negative value

Figure 9. Ali and Leo's small cookie weight model (a) and second attempt at a range of values for amount of playdough factor (b)

Leo's suggestion to include more values (Row 43) prompts Ali to remember a way to create a discrete spinner covering the range from -10 to 10, an aspect that they discovered by clicking on all the buttons below the spinner. He checks that all the values are present (Figure 10) and says, "That is good" [Sanctioning himself]. Their improved model (Figure 10) represented an important mid-stage conceptual shifting from a discrete modelling tool to a continuous tool to better represent measurement data in the model world. Their dialogue also seemed to show that comparing and contesting representations aided this conceptual shift. Note also that Leo is no longer suggesting percentage values.

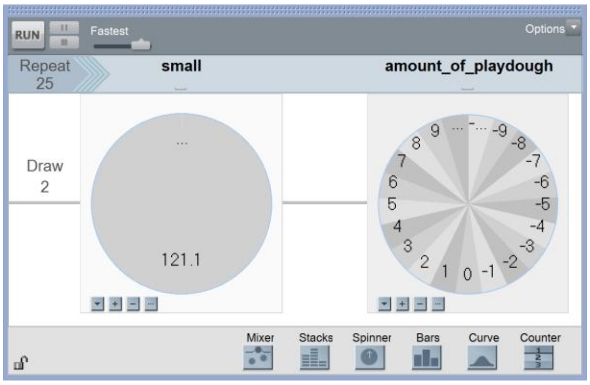


Figure 10. Ali and Leo's third attempt at a range of values for amount of playdough factor that includes all integer values between -10 and 10

Fifth episode In this episode (Table 4), the final variable, "thickness," and the final model (Figure 11) are determined. Two of Ali's utterances, Rows 48 and 51, are dual coded as conceptualising language (yellow shading) and concretising language (blue shading) because there is simultaneously a resultant action and a sanctioning of Leo's suggestions. All utterances are now classified as conceptualising language because there is an indication that the concepts are in the process of formation based on the past commitments and history of the student responses. Even though Leo only states, "by 1 to 15" in Row 50, it is coded as *giving a reason* as it makes sense to Ali, who *sanctioned* it with his *action* in Row 51. We also know it is a reasoned response based on past episodes and past utterances (e.g., Table 3, Row 35). Finally, in Rows 52 and 53, Ali and Leo *notice* and simultaneously *sanction* the outcomes (Figure 12) of their final model (Figure 11).

Table 4. IB framework coding

Row	Action	Interrogative cycle	Who	Utterance	Concept(s) in process of formation
44		Wondering	Ali:	Ok, can you think of another thing, variable?	Causes of variation
45		Wondering	Leo:	Maybe the colour of playdough?	Causes of variation
46		Censuring himself	Leo:	No.	
47		Wondering	Leo:	Oh, I know, maybe how good you roll it?	Causes of variation
48	Creates a thickness spinner	Sanctioning	Ali:	Yeah.	Causes of variation
49		Wondering	Ali:	Ok, tell me how much that's going to affect the data by Leo.	Sample space
50		Giving a reason	Leo:	By 1 to 15	Sample space
51	Enters a range from -15 to 15. Runs model several times, (Figures 12a and 12b)	Sanctioning	Ali:	Let's just run it (Figure 11) and see how it (the simulations compare to the actual data in Figure 12c) goes.	Model fit
52		Noticing Sanctioning	Ali:	Yeah, I'm happy with that.	Model fit
53		Noticing Sanctioning	Leo:	Yeah, so am I.	Model fit

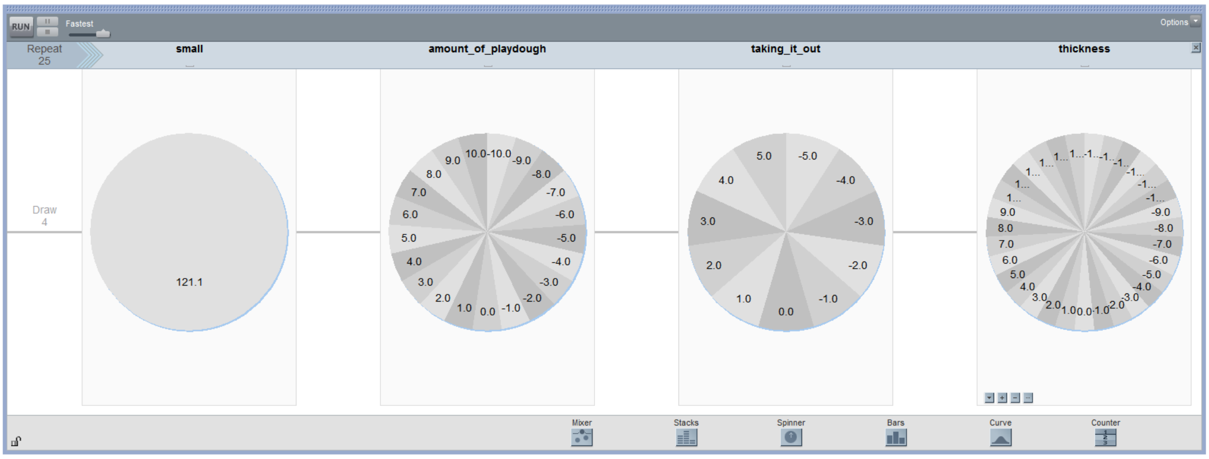
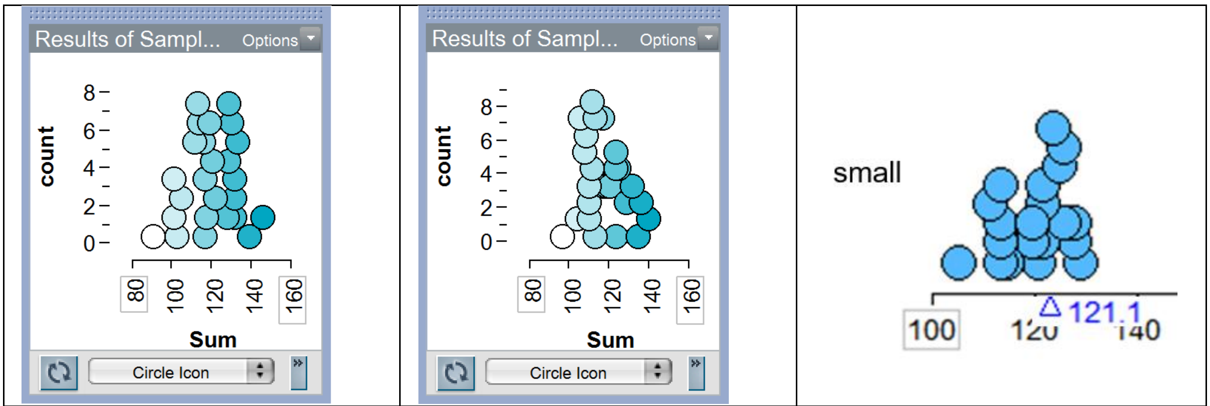


Figure 11. Ali and Leo's final model for the small cookie weights



(a) One simulation | (b) Another simulation | (c) Small cookie weight distribution

Figure 12. Two simulations of Ali and Leo's model (a) and (b) and real data distribution (c)

Note Ali understands Leo's error size suggestion of 1 to 15 to be the range -15 to 15 (Row 51). Also, note that in the third to fifth episodes how they adopted an interrogative interplay between themselves with Ali taking on the role of the "teacher." Their dialogue showed them comparing and contesting representations and comparing real data distributions to simulated distributions, highlighting how the interplay among representations, reasoning, and judging might assist conceptual formation about sample space, causes of variation, and model fit. They used the data context to determine which factors should be included to represent variation from the mean in their cookie weight model. Moreover, the data context continued to help Leo distinguish between factors that vary and factors that do not (Rows 45 and 46). Ali concretised the structure of their model that needed to be built, and Leo responded to Ali's questions by suggesting names for the variables and the size of each factor representing variation from the mean.

Together these students were negotiating shared meanings and common language about how they would go about constructing their model to replicate or mimic the real data. Over the two tasks, the researcher participated in the dialogue to help these students make sense of what they perceived and conceived about their model and with their reasoning, decision-making, and actions by drawing inferential connections between concepts and between concepts and actions.

5. DISCUSSION

The inferentialism-based (IB) framework that was developed involved interconnected inferential cycles comprising interrogative cycles of noticing and wondering (Shaughnessy, 1997), giving and asking for reasons, and sanctioning and censuring (Brandom, 2000), as well as oscillations between concretising language about actions and conceptualising language towards concept formation (Heusdens et al., 2019). As illustrated in the five episodes, the concepts encountered were navigated within interconnected inferential cycles that involved the students acting on displays, drawing on contextual aspects of the task scenario to make sense of them, and comparing and contesting them, which seemed to reshape these individuals' webbing and build on their concepts through inferential language and actions. As Konold and Kazak (2008, p. 30) stated, "seeing something new" is "synonymous with learning a new conception," and "what we know highly influences what we can see; what we see drives what we can come to know." According to our IB framework, concepts appear to form and grow through the interplay between concretising, which may or may not lead to actions (top broken arrow in Figure 1), and conceptualising, which may or may not lead to concept formation (bottom broken arrow in Figure 1). The interplay seems to be an interrogative oscillating process that may strengthen inferential webbing between concepts through noticing and wondering, the giving and asking for reasons, and the sanctioning and censuring of reasons. Thus, a conjecture is that the inferential cycles may provide a way of understanding the formation process of students' concepts.

Nilsson et al. (2018, p. 374) identified that current frameworks and research articles in statistics education required a "stronger theoretical basis or treatment" of knowledge construction and that frameworks needed to be grounded in learning theory. Building on the work of other researchers (e.g., Bakker et al., 2017; Bakker & Derry, 2011; Heusdens et al., 2016), they suggested inferentialism as a potential theoretical resource for learning about how conceptual infrastructure might be formed. Furthermore, Nilsson et al. (2018) proposed four aspects where inferentialism might provide some insights into student learning, namely concepts in use, a dynamic and holistic view of concept formation, the interplay between technology and human-based decisions and interpretation, and contextualising. We contend that our IB framework might have the capability to address these four aspects in statistics education research.

The first aspect was that inferentialism might offer a new perspective on concepts in use between individuals and others—the social sphere. By applying the framework to the data collected, it became apparent that the act of statistical modelling and the concepts in use are regulated by individuals' shared reasoning and actions. What participants noticed and attended to were both in-the-moment and over time, which influenced their inferences, and therefore which concepts were strengthened and which faded. For example, the analysis showed how they tested ideas about causes of variation, continuous measurement, and appropriate ranges of values to represent variation, resulting in the strengthening of ideas that were sanctioned and actioned and possibly the fading of ideas that were censured. Concept formation can also depend on the power relationships between participants such as the authority of the teacher and what notions are privileged.

The second aspect was that inferentialism might offer a "dynamic and holistic view" of concept formation and thus "avoid static usage of frameworks with categories or levels" (Nilsson et al., 2018, p. 379). Our framework is dynamic, capturing learning over time through continuous interrogative cycles that oscillate between concretising language for actions and conceptualising language that provides a pathway to concept formation. It is also holistic in that it narrates how concepts are encountered and navigated and recognises that concepts do not occur in isolation, rather that statistical concepts are learnt together because concepts only have meaning when related to other concepts (Bakker & Derry, 2011). For example, the students were simultaneously relating mean, sample space, and variation to each other and to the concepts of simulated and real distributions using informal sense-making novice language—a prerequisite in the learning process towards formal language and definitions (Arnold & Pfannkuch, 2024).

In particular, the concept of causes of variation grew from their contextual experiences of manufacturing cookies to realising that those experiences could be transformed into a model that they could build (cf. Konold & Harradine, 2014). The students' experiences of differences between the cookies resulted in them constructing spinners for perceived differences or causes of variation and coming to a consensus on what was accepted as a cause. In the construction of the spinners, the concept of sample space came into play, whereby students contested the size of the variation that could be contributed by the factor, including how many values to use. Agreeing on what constituted an acceptable sample space for each factor also involved drawing on the concept of an acceptable sample space for the simulated distributions and hence became part of the concept formation process. The concept of model fit began from the first episode for the distribution of moon hopper circumferences, when the researcher drew students' attention to what was an acceptable range of values and to distributional shape and central tendency. Over time, the students came to a consensus on what they perceived as an adequate match between the real data distribution and the simulated distributions, that is, they appeared to be forming the concept of model fit (cf. Konold & Kazak, 2008).

Dvir and Ben-Zvi's (2018) contention that students' acts of comparison were important in learning statistical concepts seemed to be affirmed through the use of the IB framework. Research (e.g., Konold & Harradine, 2014) has shown how these concepts may be constructed within a technology-based statistical modelling environment but has not used such a fine-grained account to track the process of conceptual formation (cf. Derry, 2017). Thus, the analysis allowed a window into how concepts may grow over time, from tentative suggestions to a growing confidence in the application of the concepts, as participants and the technology reinforced the acceptability of their use (cf. Bakker & Derry, 2011).

The third aspect of inferentialism was related to technology, which Nilsson et al. (2018) argued might provide a window into the interplay between technology and human reasoning and what educators might need to attend to when compared to a pencil and paper learning environment. Ali and Leo used a trial and error approach based on their contextual knowledge of the situation to create their model, spinner attributes, and range of values, each time testing each additional spinner against the real data distribution for model fit. Technology lends itself to a trial and error approach, a well-known instinctive problem solving strategy that humans adopt when confronted with a new situation. Because it is so quick to trial and test possible ideas using technology, educators may use that capability as part of the task design for learning new ideas (cf. Fergusson & Pfannkuch, 2022) or as a warning about the need to ensure that students reflect on other problem solving strategies. These strategies can include the development of other knowledge and concepts that students can draw on or learn and what counts as a different, sophisticated, efficient, or acceptable strategy or one that leads to generalisation (cf. Noll et al., 2018). The technology also facilitated and coordinated student actions and reasoning in the *TinkerPlots* modelling environment. For example, Ali and Leo clicked on all the buttons below the spinner and discovered that they could enter a range of numbers rather than typing a number into each of the three spinner divisions as they had been taught. Hence, they used the technology to better represent variation in their model that matched their reasoning. Effectively they produced pseudo-continuous data using a discrete spinner by using the range function on the spinner tool, which signalled an important mid-stage conceptual shift from a discrete modelling tool to a continuous tool to better represent measurement data in the model world. Thus, technology seemed to assist these students to infer connections between the tool and their reasoning, resulting in more sophisticated nascent concepts of variation and measurement error. Technology provided a scaffolded environment that enabled these students to visualise and test their inferences about the content and application of concepts as they strengthened and developed.

The fourth aspect was that inferentialism might shed light on how students "learn to contextualise and integrate statistical and contextual considerations" (Nilsson et al., 2018, p. 380). Our analysis of the student dialogue showed many examples of students using their experiences of the data context to suggest factors that might contribute to variation. Hence, they were learning to contextualise through the interrogative cycles. When they were considering what factors would vary, they were learning to integrate the statistical with the contextual. For example, peer censuring from Ali enabled Leo to self-censure later on when he suggested cookie colour as a factor for the model they were building. In all their public interrogative discussions, the students' reasoning privileged the data context, which appeared to have been used as a referent to justify their actions and interpretation of the representations. Lesh and Doerr (2003) pointed out that a critical aspect of modelling a realistic situation was for students to make sense of the situation based on extensions of their own personal knowledge and experiences. Moreover, the data context appeared to support inferential connections between their prior modelling experiences in the Moon Hopper Task and the Cookie Cutter Task regarding the error structure required to model both situations. Additionally, it provided coherence in the students' public reasoning about features seen in real and simulated data and possible factors underlying real data distributions. Therefore, our framework has the potential to provide insights into student learning regarding the four aspects that Nilsson et al. (2018) thought the theory of inferentialism might be able to address.

Using the IB framework as an analysis tool enhanced the researcher's awareness of the teacher's role in inducting and promoting the students' interrogative interplay of noticing and wondering, the giving and asking for reasons, and sanctioning and censuring. Her modelling of the interrogative interplay (cf. Derry, 2017) may have encouraged Ali and Leo to adopt it between themselves, which also might encourage students to learn that articulations about their models are an inherently normative practice. The IB framework seemed to have the capability of identifying students' emerging statistical concepts and examining how individuals coordinate their unique conceptions and actions through strengthening inferential webbing. As a result of creating and using the IB framework to explicate students' reasoning over time, the findings could affirm Radford's (2017, p. 505) view that the reformulation of inferentialism could offer "a fresh perspective on knowledge, concept formation, and learning that privileges inferential thinking" as well as offering "the ability to think in new ways about the question of task design and pedagogical action." Thus, the IB framework contributes to the debate within statistics education as to the viability of inferentialism as a learning theory.

However, Radford's (2017, p. 506) critique of inferentialism is whether it really addresses the question of "what makes us human," as he would rather focus on the capacity of humans to live with and respond to each other with empathy and care. Although our IB framework could not address Radford's concerns, we conjecture that it could, within an expanded version, show how empathy and social awareness could be activated through the context of the investigation. For example, stimulated by media reports citing research articles, the students collected and modelled data from students in their class on backpack weights and weekly homework hours, which may have begun to produce awareness about societal debates about back problems caused by heavy schoolbags and excessive homework being set for students. Thus, future research using this data might be able to show students becoming socially and politically aware. The IB framework was trialled on data gathered from a small group of students over 12 two-hour lessons in a two-year period. The exploratory nature of the study means that the framework is in its infancy and needs to be extensively used in future research to understand whether inferentialism theory can indeed be used in practice and whether it can provide insights into student learning and concept formation that could lead to improvements in statistics pedagogy.

ACKNOWLEDGEMENTS

This article is a substantial elaboration of our six-page paper published in the proceedings of the International Conference on Teaching Statistics (ICOTS-11), *Bridging the gap: Empowering and educating today's learners in statistics* (Patel & Pfannkuch, 2022). We thank the reviewers for their constructive feedback and very helpful suggestions on how to improve our paper.

REFERENCES

Arnold, P., & Pfannkuch, M. (2024). Engaging novice statisticians in statistical communications. *Mathematics Education Research Journal*, *36*(Suppl 1), 147–173. https://doi.org/10.1007/s13394-022-00442-w

Bakker, A., Ben–Zvi, D., & Makar, K., (2017). An inferentialist perspective on the coordination of actions and reasons involved in making a statistical inference. *Mathematics Education Research Journal*, *29*(4), 455–470. https://doi.org/10.1007/s13394-016-0187-x

Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, *13*(1–2), 5–26. https://doi.org/10.1080/10986065.2011.538293

Bakker, A., & Hußmann, S. (2017). Inferentialism in mathematics education: Introduction to a special issue. *Mathematics Education Research Journal*, *29*(4), 395–401. https://doi.org/10.1007/s13394-017-0224-4

Baroody, A. J., Cibulskis, M., Lai, M., & Li, X. (2004). Comments on the use of learning trajectories in curriculum development and research. *Mathematical Thinking and Learning*, *6*(2), 227–260. https://doi.org/10.1207/s15327833mtl0602_8

Brandom, R. (2000). *Articulating reasons: An introduction to inferentialism*. Harvard University Press. https://doi.org/10.4159/9780674028739

Charmaz, K. (2014). *Constructing grounded theory*. SAGE.

Chernoff, E., & Zazkis, R. (2011). From personal to conventional probabilities: From sample set to sample space. *Educational Studies in Mathematics*, *77*(1), 15–33. https://doi.org/10.1007/s10649-010-9288-8

Derry, J. (2017). An introduction to inferentialism in mathematics education. *Mathematics Education Research Journal*, *29*(4), 403–418. https://doi.org/10.1007/s13394-017-0193-7

Dvir, M., & Ben-Zvi, D. (2018). The role of model comparison in young learners' reasoning with statistical models and modeling. *ZDM–Mathematics Education*, *50*(7), 1183–1196. https://doi.org/10.1007/s11858-018-0987-4

Fergusson, A., & Pfannkuch, M. (2022). Introducing high school teachers to predictive modelling and APIs using code-driven tools. *Statistics Education Research Journal*, *21*(2), Article 8. https://doi.org/10.52041/serj.v21i2.49

Heusdens, W., Baartman, L., & De Bruijn, E. (2019). Know your onions: An exploration of how students develop vocational knowledge during professional performance. *Scandinavian Journal of Educational Research*, *63*(6), 839–852. https://doi.org/10.1080/00313831.2018.1452291

Heusdens, W., Bakker, A., Baartman, L., & De Bruijn, E. (2016). Contextualising vocational knowledge: A theoretical framework and illustrations from culinary education. *Vocations and Learning*, *9*(2), 151–165. https://doi.org/10.1007/s12186-015-9145-0

Kaplan J., & Peters, S. (2023). *Publishing in SERJ: Processes and suggestions* [Webinar]. International Association for Statistical Education. https://www.youtube.com/watch?v=rhMig3ZhZBU

Konold, C., & Harradine, A. (2014). Contexts for highlighting signal and noise. In T. Wassong, D. Frischemeier, P. Fischer, R. Hochmuth, & P. Bender (Eds.), *Mit Werkzeugen Mathematik und Stochastik lernen: Using tools for learning mathematics and statistics* (pp. 237–250). Springer. https://doi.org/10.1007/978-3-658-03104-6_18

Konold, C., & Kazak, S. (2008). Reconnecting data and chance. *Technology Innovations in Statistics Education*, *2*(1), 1–37. https://doi.org/10.5070/T521000032

Konold, C., & Miller, C. (2011). TinkerPlots: Dynamic data exploration (Version 2.3) [Computer software]. 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, July 2015)* (pp. 76–86). University of Paderborn.

Lehrer, R., & English, L. (2018). Introducing children to modeling variability. In D. Ben–Zvi, K. Makar & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 229–260). Springer. https://doi.org/10.1007/978-3-319-66195-7_7

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

Noll, J., Clement, K., Dolor, J., Kirin, D., & Petersen, M. (2018). Students' use of narrative when constructing statistical models in TinkerPlots. *ZDM–Mathematics Education*, *50*(7), 1267–1280. https://doi.org/10.1007/s11858-018-0981-x

Noorloos, R., Taylor, S., Bakker, A., & Derry, J. (2017). Inferentialism as an alternative to socio–constructivism in mathematics education. *Mathematics Education Research Journal*, *29*(4), 437–453. https://doi.org/10.1007/s13394-017-0189-3

Nilsson, P. (2020). Students' informal hypothesis testing in a probability context with concrete random generators. *Statistics Education Research Journal*, *19*(3), 53–73. https://doi.org/10.52041/serj.v19i3.56

Nilsson, P., Schindler, M., & Bakker, A. (2018). The nature and use of theories in statistics education. In D. Ben–Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 359–386). Springer. https://doi.org/10.1007/978-3-319-66195-7_11

Patel, A. (2022). *Statistical modelling: An enquiry into novice students' co-creation of reasoning and practice.* [Unpublished doctoral dissertation]. The University of Auckland. https://hdl.handle.net/2292/58778

Patel, A., & Pfannkuch, M. (2018). Developing a statistical modeling framework to characterize year 7 students' reasoning. *ZDM–Mathematics Education*, *50*(7), 1197–1212. https://doi.org/10.1007/s11858-018-0960-2

Patel, A., & Pfannkuch, M. (2022). A framework for capturing the development of statistical concepts. In S. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the gap: Empowering and educating today's learners in statistics. Proceedings of the Eleventh International Conference on Teaching Statistics (ICOTS11, September 2022), Rosario, Argentina.* International Association for Statistical Education. https://doi.org/10.52041/iase.icots11.T2A1

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

Radford, L. (2017). On inferentialism. *Mathematics Education Research Journal*, *29*(4), 493–508. https://doi.org/10.1007/s13394-017-0225-3

Radford, L. (2018). On theories in mathematics education and their conceptual differences. In B. Sirakov, P. de Souza, & M. Viana (Eds.), *Proceedings of the International Congress of Mathematicians,* (Vol 4. pp. 4055–4074). World Scientific Publishing. https://doi.org/10.1142/9789813272880_0211

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

Taylor, S., Noorloos, R., & Bakker, A. (2017). Mastering as an inferentialist alternative to the acquisition and participation metaphors for learning. *Journal of Philosophy of Education*, *51*(4), 769–784.

Van Oers, B. (1998). From context to contextualizing. *Learning and Instruction*, *8*(6), 473–488. https://doi.org/10.1016/S0959-4752(98)00031-0

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning.* The Falmer Press

ANNE PATEL
Te Kura Tatauranga | Department of Statistics
Waipapa Taumata Rau | The University of Auckland
Private Bag 92019
Tamaki Mākaurau Auckland 1142
Aotearoa New Zealand