

Educ Stud Math  
DOI 10.1007/s10649-013-9526-y

---

## Young children's explorations of average through informal inferential reasoning

Katie Makar

© Springer Science+Business Media Dordrecht 2013

**Abstract** This study situates children's early notions of average within an inquiry classroom to investigate the rich inferential reasoning that young children drew on to make sense of the questions: *Is there a typical height for a student in year 3? If so, what is it?* Based on their deliberations over several lessons, students' ideas about average and typicality evolved as meaning *reasonable*, contrary to *atypical*, *most common* (value or interval), *middle*, *normative*, and *representative of the population*. The case study reported here documents a new direction for the development of children's conceptions of average in a classroom designed to elicit their informal inferential reasoning about data.

**Keywords** Average · Statistics education · Informal inferential reasoning · Informal statistical inference · Inquiry-based learning

Average is a central concept in statistics. While most older students can calculate an arithmetic mean, nearly all struggle to use it meaningfully (Konold & Pollatsek 2002; Lavigne & Lajoie 2007). Bakker and Derry (2011) argue that this inert knowledge is a persistent challenge in statistics education that flows from an endurance of atomistic approaches to learning statistics that for students, lack coherence (see descriptions of “meanmedianmode” phenomena in Bakker 2004 & Friel et al. 2006). Rather, there is a need to embed learning statistics holistically within processes of thinking and reasoning that *value the utility* of statistical tools rather than only *knowledge* of them (Ainley, Pratt, & Nardi 2001; Bakker & Derry 2011). To combat an atomistic approach, Bakker and Derry recommend a focus on inferentialism, which “puts inference [and reasoning] at the heart of human knowing” (p. 6). Acknowledging that the primary reason why people use statistics is to make inferences about a population or process based on available data, a new research perspective reframes statistical inference more broadly as an uncertain claim beyond available data (Makar & Rubin 2009).

Benefits for young children developing informal inferential reasoning have been explored (Ben-Zvi, Gil, & Apel 2007; Paparistodemou & Meletiou-Mavrotheris 2008), acknowledging that

---

Some data in this paper were previously presented at the 32nd Annual Conference of the Mathematics Education Research Group of Australasia. This paper significantly extends the data, analysis, theoretical underpinnings and emphasis on informal inferential reasoning.

K. Makar (✉)  
School of Education, The University of Queensland, Brisbane 4072, Australia  
e-mail: k.makar@uq.edu.au

repeated and long-term exposure could enable richer statistical understandings and support learning when formal ideas are later introduced (Ben-Zvi et al. 2007; Makar, Bakker, & Ben-Zvi 2011). Using rich classroom contexts in research which operationalise statistical learning parallels a broader shift towards more process-oriented educational research that both demonstrates an appreciation for mathematical practices developed through repeated experiences and documents how students *build* mathematical ideas rather than documenting current or changes in reasoning (Harel & Koichu 2010; McGowen & Tall 2010; Mercer 2010; Scardamalia & Bereiter 2006).

This paper reports on a year 3 (age 8) class as they wrestle with the questions, *Is there a typical height for a person in year 3? If so, what is it?* The study extends previous research on children's conceptions of average by (1) documenting its *emergence* in an inquiry-based classroom over several weeks, (2) prioritising the building of concepts using *informal statistical inference* over descriptive statistics, and (3) capitalising on *ambiguity of language* ("typical") to provoke student negotiation of multiple meanings of average. Previous research on students' understanding of average has primarily been carried out through individual interviews (e.g., Mokros & Russell 1995; Watson 2007), with a few students (e.g., Lavigne & Lajoie 2007) or in short-term teaching experiments taught and/or designed by researchers (e.g., Cobb 1999; Lavigne & Lajoie 2007). In addition, most research has focused on students' (current) understanding or change in understanding of average as part of descriptive statistics (e.g., Watson 2007). By considering students' early development of informal conceptions of "typical" in a classroom that promotes informal inferential reasoning and student negotiation, the data provide insight into how students' understanding of complex statistical ideas can be initiated from an early age and illustrate an operationalisation of a more holistic, process-oriented approach to learning statistics (Bakker & Derry 2011).

## 1 Informal inferential reasoning

The power of statistics lies in making inferences about the world beyond available data. Statistical inference taught at university focuses on the techniques accepted by the discipline to make claims from a properly collected sample to a larger population. Informal statistical inference is a broader concept designed to introduce those less experienced in statistics to inferential processes without the formal methodologies (Pfannkuch 2006; Rubin, Hammerman, & Konold 2006). Briefly, an informal statistical inference uses available data as evidence to make an uncertain claim about the world and is characterised as having the following three key components: a *generalisation* or claim that extends beyond the data, use of data as *evidence* to support the claim, and articulation of uncertainty through *non-deterministic language* (Makar & Rubin 2009).

Researchers recognise the powerful influence that an inferential approach can have on children's understanding of statistics (Bakker & Derry 2011; Papanistodemou & Meletiou-Mavrotheris 2008; Pratt & Ainley 2008). There are concerns, however, that by introducing it into school curriculum, informal statistical inference may be reified to become the next "meanmedianmode" (Bakker 2004; Friel et al. 2006). That is, that informal statistical inference will be taught to students as an entity in and of itself, rather than focusing on the reasoning used to draw useful context-rich conclusions about data. To guard against this, the focus must be on the reasoning that leads to inference rather than the statistical inference itself (a statement). The following characteristics have been argued to support informal inferential reasoning:

- An *inquiry-based* environment which builds students' collaborative norms
- *Statistical concepts and tools* to support and extend the construction of informal statistical inferences from the data

- Data-rich tasks that trigger *conflicts with beliefs* (contextual and/or statistical) to encourage students to seek deeper insights and explanations (Makar et al. 2011).

A classroom culture which promotes inferential reasoning requires that students address problems in an environment which encourages collaborative norms, statistical concepts (formal or informal) and a sufficiently complex problem context to enable them to encounter conflicts with their knowledge and beliefs about the world. Therefore, students need more than the clean, pre-designed investigations that they typically encounter in school statistics; rather, they need opportunities to work with the messy, incomplete data collected from authentic contexts (Gould 2011; Rubin 2005). Context adds a level of ambiguity and depth to problems that requires different skills from school-type problems and enables learners to make connections that link statistical measures and methods with the worldly problems that they were designed to investigate (Gal 2002). The difficulty is in assisting students to *coordinate* statistical and contextual understandings to reconcile the gap between what they know from experience and what they learn from data (Wild & Pfannkuch 1999). New recommendations emphasise that focusing school statistics on informal statistical inference provides a natural opportunity to integrate context into statistical learning (Makar & Ben-Zvi 2011; Pratt & Ainley 2008).

## 2 The concept of average

Although most students can identify the formula for the mean, few use it appropriately to solve problems (Lavigne & Lajoie 2007; Shaughnessy 2007). “What has been striking over 25 years of research, is the difficulty encountered by students of all ages in coping with the representative nature of the arithmetic mean” (Watson 2007, p. 22). In their classic study, Mokros and Russell (1995) identified five conceptions of average held by children (aged 9–14 years old) as follows: mode, algorithm, reasonable, midpoint and as a mathematical balance point. Of those who relied on the algorithm, none used it effectively and gave the algorithm absolute credence even when reasonableness and self-monitoring indicated a different answer. Transfer advocates contend that over-development of “verbatim memory” in isolation creates a division between memorisation and reasoning, rendering the latter inert and impoverished (Wolfe, Reyna, & Brainerd 2005). In this light, Mokros and Russell's students' neglect of engaging sense-making mechanisms over their rehearsed algorithmic approach to average is unsurprising.

To reconceptualise average, researchers have advocated approaches which portray average as representative of an aggregate. Concepts of centre such as “modal clump” (Konold et al. 2002) or “signal and noise” (Konold & Pollatsek 2002) take into account most or all of the data. However, most conceptions of average articulated by students in Mokros and Russell's (1995) study prioritise algorithmic conceptions or neglect an understanding of average as representative of a distribution. For example, mode or midpoint views tend to focus students on a single value, rather than conceptualising data as an aggregate (Konold, Higgins, Russel, & Khali 2004). Average as a balance point or “fair share” has been argued as a proxy for the algorithmic approach (Ginat & Wolfson 2002; Lampen 2013).

To combat problems with the children's difficulty with average, Mokros and Russell suggest that children repeatedly encounter informal notions of average within authentic contexts “to develop their own ideas of typicality” (p. 38) and provoke notions of reasonableness. They contend that by age 9,

Students have developed powerful, situation-based ways of thinking about average ... Students' notions of representativeness or typicality grow out of everyday experiences and have a strong flavor of reasonableness and practicality ... Children's informal ideas

about outliers helped them home in on what was typical. Reasonableness in evaluating a data set appeared as an essential strand in understanding, a strand that plays a significant role in the development of more complex notions of average. (p. 21)

Their study suggests several key ideas that support student learning about average as follows: (1) young children's informal notions of average are *situation-based*, (2) these notions are also grounded in their everyday experiences with *typicality* and *representativeness*, (3) experiences with *outliers*, or atypical data help them focus on what is typical, and (4) essential to more complex notions of average is the idea of *reasonableness*. What appears important here is to attend first to building learners' broader conceptions of average, assisting them to connect its diversity and complexity to experiences and situations in which typicality, representativeness and reasonableness have meaning.

Watson strongly recommends that average be developed slowly over the course of schooling with formal instruction delayed until students had more robust informal understandings. She asked students to consider average in various contexts to elicit their informal, algorithmic and conceptual understandings. By showing video-taped responses from students that differed from their own response, she used cognitive conflict to prompt students to reconsider and then re-articulate their understandings. In doing so, students were able to discuss concepts with more depth and complexity (Watson 2007). Building conceptions through experiences that promote negotiation and inquiry are enhanced by including some level of ambiguity to promote "interpretive flexibility" (p. 377) which gives students opportunities to wrestle with the multiple meanings (Roth 1995).

This paper reports on students debating and collectively wrestling with broad notions of average in an inquiry-based classroom which developed informal inferential reasoning. The ambiguity of the word "typical" was used to explore and negotiate multiple meanings of average. Drawing on the research above, a conceptual framework of average was employed to develop more productive conceptions of average in an inquiry-based classroom context.

### 3 Conceptual framework

Statistical inquiry is a pedagogical process in which a teacher supports students as they wrestle with ill-structured data-based problems (Makar & Fielding-Wells 2011). In ill-structured problems, the problem statement or solution pathways have a number of ambiguities which require negotiation to mathematise and structure the solution process (Reitman 1965). Inquiry requires an epistemological foundation that presumes knowledge is socially constructed through collaborative and iterative cycles of investigation and debate. The assumption was that data studied in an authentic environment and one which asked students to make inferences would produce deeper levels of understanding (Dierdorff, Bakker, Eijkelhof, & van Maanen 2011). There was an effort, therefore, to encourage students to consider questions such as, "What would you expect the typical height to be in the class next door?" or "What would you predict the height might be for a new student entering our class?" These questions prompted students to go beyond description of their data towards making generalisations evidenced by the data they had collected.

Watson (2006) contends that average is closely linked to concepts of inference in that one needs to think about typicality and representativeness in order to infer. Drawing on literature on students' conceptions of average, informal inference and inquiry, four key concepts were seen as relevant to support young children's explorations of average in the context of investigating students' typical heights in an inquiry classroom. These elements assisted in planning the learning environment and sensitising our observations and discussions in order to promote students' developing understandings.

- Reasonableness: To conceptualise average, students need to build a sense of *reasonableness* (Mokros & Russell 1995). By working with and measuring their own heights, the hope was that students would develop a sense of values that were reasonable in this context.
- Outliers: In deliberating on what are average or typical heights, students may find it useful to make sense of *outliers*, or values that are atypical (Mokros & Russell 1995). The assumption was that asking students to consider atypical heights, would make the consideration of typical more apparent.
- Typical as most common: Mode is a fairly conventional use of average at the primary level (Watson 2006); students often consider values with the highest frequency as being typical. There was an assumption that students at this age would consider *most common* as one definition of typical.
- Comparing groups: Group comparisons can provide the impetus for inferential reasoning (Watson & Moritz 1999). If the typical value is not the same for two classes, informal notions of sampling variability may emerge, a critical concept in informal statistical inference (Ben-Zvi et al. 2007).

Researchers (e.g., Ainley et al. 2001; Mokros & Russell 1995) have noted that students are often drawn to the middle of a dataset (midpoint or middle value) to express typicality. This was a concept that was anticipated as possibly emerging; however, it was not planned as part of the learning trajectory as it was unclear whether it would lead to more productive notions of typicality. Notably, the assumption that students would see typical as representative was also not part of the learning context, although this is arguably one of the most important conceptions of average that students would need to develop (Watson 2007). Given that (1) this is a very difficult concept even for adults and (2) it was not anticipated that this issue would arise in this particular problem context, this concept was not part of the hypothesised learning trajectory for these young children.

## 4 Context and method

The research question was, *How do young children conceptualise “average” in an inquiry-based learning environment that develops informal inferential reasoning?* To address this, a series of lessons was analysed from a year 3 classroom (26 children, age 8) in a large suburban school in Queensland, Australia in which children had been immersed in an inquiry culture.

### 4.1 Design

The research reported here is situated in a longitudinal study (2006–2012) into teachers' emerging practises in teaching inquiry in mathematics and statistics (Makar 2011; Makar & Fielding-Wells 2011). The study (both the larger study and this case study) used a Design Research framework (Cobb, Confrey, Disessa, Lehrer, & Schauble 2003) in which the researcher simultaneously studies and works to improve the context under investigation. The teacher was a participant in the longitudinal study who had been teaching mathematical inquiry and informal inference with the support of the researcher for a year. She was a highly experienced teacher with specialisations in literacy and art. She was particularly skilled at developing children's reasoning in general, but was new to the concept of inquiry and inference in teaching mathematics and statistics. The researcher and teacher met after each session to discuss their observations and make recommendations about the direction of the next lesson. Final decisions were left to

the teacher, who designed the unit, as the larger study aimed at supporting and studying teachers' evolving inquiry pedagogy, but not directing it.

## 4.2 Curriculum trajectory

The lessons took place in the final (term 4) inquiry unit of the year. Students had already completed three data-based inquiry units, each lasting 2–4 weeks. In each unit, students were asked to make predictions and inferences about contexts beyond their classroom. The previous three units were as follows:

- *Investigating our hand spans* (term 1)
- *Do we eat a healthy lunch?* (term 2)
- *What kinds of appliances do we have at home?* (term 3)

The classroom culture was one in which the students repeatedly shared and discussed emerging ideas and were encouraged to debate and articulate their reasonings as they evolved. These skills were developed not only during the data-based inquiry units, but also more broadly. The inquiry units, however, were distinct from the teacher's conventional mathematics lessons which relied more heavily on a prescribed textbook.

### 4.2.1 Previous units (terms 1–3)

In *Hand Span*, students worked with their class hand span data to learn to describe data, compare data from a neighbouring class, and make predictions about other year 3 classes (see Makar & Rubin 2009 for more detail). Much of the learning was addressed through students' diverse attempts to collect, record and interpret the data, with the class pausing frequently to share progress, problematise issues, support those who were struggling to make headway, debate potential alternatives and propose efficiencies (Allmond, Wells, & Makar 2010). Students eventually made inferential predictions about the hand spans of year 3 students as a range of "about 15–17 cm". In the second and third units of the year, students collected, organised and analysed increasingly complex sets of categorical data. These units integrated their mathematics learning with work in other content areas (health, social studies, science) and gave students extended experiences with negotiating possible categories of food or appliances and managing larger datasets.

### 4.2.2 Typical heights (term 4)

In the unit described in this paper, students investigated the driving questions: *Is there a typical height for year 3 students? If so, what is it?* This task revisited some of the concepts introduced in the Hand Span unit (term 1) as well as extending a much greater range of measurements that provoked more complex issues of distribution, sampling variability and inference. In the typical height unit, students proposed that they measure heights of the students in the class, came up with a plan for collecting and categorising the data and noted salient features of their data to predict and compare their class data with heights in the neighbouring class. The unit lasted for 3½ weeks, with approximately 3–4 lessons per week lasting 60–90 min each. The general trajectory was sketched ahead of time, but students' ideas were the impetus for the direction of individual lessons under the skilful guidance of the teacher. That is, key activities were the result of class discussions which had identified, through student input and/or teacher prompting, the need for them. The weekly focus is summarised in Table 1.

**Table 1** Overview of key activities videotaped in the typical height unit

Week	Key activities (data collection dates)	Emerging conceptions of typical (average) height
1	Discuss inquiry question and suggest a plan; measure class, establish measurement and data collection protocols, make preliminary observations (25, 29 Oct)	Reasonable; most common (modal, single value)
2	Organise data and negotiate preliminary estimates of typical height (5 Nov)	Most common (interval)
3	Collect data from the neighbouring class; organise new data and compare with their own data; make initial inferences (9, 12, 15 Nov)	Middle (mid-range), medium (normative, modal clump)
4	Wrap up and articulate a conclusion to the inquiry question (21 Nov)	Representative

### 4.3 Data collection and analysis

Seven lessons were videotaped during the unit. Not all lessons were videotaped, for example, when students were primarily measuring heights or collecting data. The video data were analysed using methods adapted from Flick (2009). Video logs were created for each lesson to provide a descriptive sequence of activities and to locate broad excerpts requiring further analysis. Video logs were annotated by the researcher and teacher individually to provide an overall content analysis and summary of key concepts and activities in the data. Annotated notes were discussed and combined with particular attention to the conceptual framework as an initial guide. Salient episodes were transcribed for more detailed analysis, selected because of their potential to illustrate particular concepts or provide greater depth of insight about students' evolving understandings. Selected transcribed segments were then interpreted for meanings, focusing on progressions, relationships between concepts, origins and emergence of students' ideas, negative cases and explanations for the interpretations over the series of lessons. The data were scanned again to seek further opportunities for insight and new excerpts that might further deepen or contextualise interpretations. Although this process appears linear, in practice, the phases of analysis frequently included back-tracking and re-starting as understandings and interpretations shifted.

## 5 Results

Five key concepts of “typical” (average) emerged as students debated the inquiry question. Typical as (1) a reasonable height, (2) the most common value or interval of data in the class, (3) the middle height, (4) the medium (normative) population height and (5) representative of a subpopulation. Below, the progression of students' reasoning over the course of the unit is described within each category, illustrating these ideas with excerpts from student discussion.

### 5.1 Typical as reasonable

As students measured each other, initial ideas of typical encompassed values that seemed *reasonable*, arising in the following three different situations as they worked: errors in recording and measuring, extreme values (tallest and shortest people) and measurements

outside of an expected range. In each case, considering reasonableness encouraged students to check if measurements made sense. Students at this age have little experience with measurement beyond 100 cm; therefore, there were often errors in measuring and recording heights. A common error was to neglect the initial metre and only write down values beyond 100 cm (writing 35 cm instead of 135 cm). For example, Heather responded that a peer's reported height did not make sense "because 32 is a bit too short" [25 Oct, 25:42]. Her comment suggests a conflict between her expectations of reasonable heights (in relation to other measurements) and 32 cm. Instances like this gave the teacher opportunities to encourage students to develop measurement protocols (e.g., measuring students standing rather than lying down, sticking measuring tapes to the wall to avoid slack). These measurement protocols then became part of the norms of the class which students valued and monitored throughout and beyond the unit. This pedagogical approach is typical in developing inferential reasoning—utilising the emergence of conflicts and errors to drive negotiation towards new understandings (Makar et al. 2011).

Outliers triggered another discussion of reasonableness (Mokros & Russell 1995). Charles was significantly taller (155 cm) than the rest of the class, eliciting a discussion about whether Charles could represent a typical height in the class. Barbara used the word *sensible* to contrast Charles' height with those she considered typical.

*Teacher:* So are you saying here, that you do not expect a lot of people to be 155?

*Barbara:* I will show you, this. I am extra super, duper, duper sure that no one is 155 except for Charles.

*Teacher:* Because?

*Barbara:* Charles is the tallest in the class!

*Teacher:* Well what about one of the other numbers, then, 132?

*Barbara:* Well, 132 is a sensible one. I think lots of, most people will go on 132. Unless I find a more sensible one, like 130 or something. [29 Oct, 26:55]

The contrast with Charles' height shifted momentum towards values aligning with expectations of typical height. As students continued data collection, they noticed that measurements tended to fall within a fairly limited range. In this observation, they were tacitly developing a sense of reasonable values. When measurements arose outside that implicit range, it triggered a conflict.

*Patrice:* I do not think-, Barbara wrote down herself actually for 108 cm.

*Teacher:* And what are you saying about that, Patrice?

*Patrice:* She is probably not 108 cm. [5 Nov, 8:27]

These self-checking mechanisms—comparing heights with what was deemed *reasonable* or *sensible*—were triggered increasingly quickly as students concluded their data collection.

## 5.2 Typical as most common

Once students began organising the data, they noticed frequencies in their distribution. This triggered a debate of whether *typical* could mean *most common*. Students explored "most common" in their *Hand Spans* unit earlier in the year, but the current unit raised new issues. Height data were more complex because they spanned a greater range than the hand span data (with measurements covering only a few centimetres). As students organised their data, some groups ordered and tallied the heights. Two categories of typical emerged as follows: as a range of values and as the single most common value. Students initially argued "most common" as range of heights, but struggled to articulate how they defined their boundaries.



*Teacher:* [To the class] Is there a typical height for a person in year 3? ... Have you found that every person in the class is the same height? What are you finding out? Yes, Brett?

*Brett:* Well, I would think that. Um. The height of. A typical height for a person in year 3 would be like around 128 to 135.

*Teacher:* ... Why are you saying those two numbers?

*Brett:* Because it's. Um. 128 is going. Keeps on going until [long pause].

*Teacher:* Where did you get this number from, 128 cm? [long pause] Is it to do with one of the measurements that you've got in your work?

*Brett:* Yes.

*Teacher:* Have you got that somewhere in your book? [long pause] Just wondering where you got that measurement from, Brett. Did you measure somebody?

*Brett:* Well, I measured that Kim was 129 which is near 28.

*Teacher:* So somebody's 129. Did you have anybody shorter than 129?

*Brett:* Barbara.

*Teacher:* And how tall is she?

*Brett:* 120.

*Barbara:* Actually, [I'm] 122.

*Teacher:* So you're deciding that 120 couldn't possibly be within the typical range.

*Brett:* Yes.

*Teacher:* Would you like to talk more about that?

*Brett:* No. [29 Oct, 37:58]

Brett hypothesises a range of possible values for the typical height of a year 3 student that appeared to exclude the upper and lower extreme heights, but he had difficulty explaining how his tentative assertion was determined. Melanie also speculated that the typical height would be an interval, and appeared to focus on intervals that she and her peers began tallying (Fig. 1), a concept that had possibly transferred from recent work on place value.

*Melanie:* Um, well, I think is that, um, that, um, most of the, um, year 3s in the class, um, end up at, yeah, 130-something.

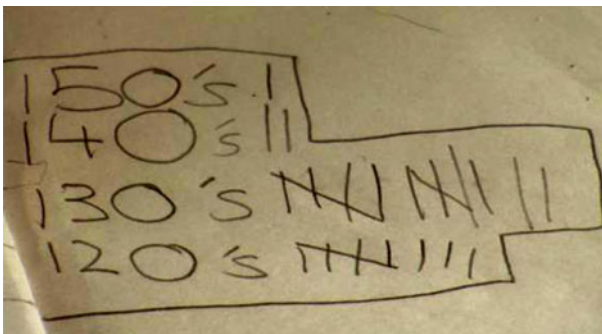
*Teacher:* ... Did you have some people that were 130-something, Melanie?

*Melanie:* Yes. But um. [long pause]

*Teacher:* Did you have lots of people? Or all of the people?

*Melanie:* Well, not all of the people. But, yes, I think most of the people would be about 130-something. [39:22]

Typical as most common was raised by Amy and Barbara as a single modal value.



**Fig. 1** Student's organisation of heights grouped by tens

*Amy:* I think typical means, like, the most popular.

*Barbara:* ... I'd actually say [the typical height is] 132. Because that's the only height, according to mine, that's [three people]...the others only have one or two. [5 Nov, 25:50]

At this stage, students put forth two main concepts for typical (1) as a range of *reasonable* values based on data collected and expectations about height (Barbara, Brett) and (2) as the single *most common* value or bin of data (Melanie, Amy, Barbara). These competing notions of typical continued in parallel for some time.

### 5.3 Typical in the population: making informal inferences beyond the data

In the excerpts above, students were using class data. Their conceptions were challenged when the teacher asked if their heights would be the same as those in the neighbouring class. This led to the following three conceptions of typical that were more inferential: typical within the population (as a general concept), typical as medium (normative) population height and typical as representative of particular populations. When Sophie proposed taking a 132-cm tall student (the most common height) from their class and “comparing” her to students next door, Elaine worried what would happen if the other class' most common height was not 132 cm.

*Elaine:* [to Sophie] What you are saying is kind of true, [but] 32 might not be the typical height. ... For example, if there were five people that were 132 cm [in our class] and there were five-, six people in [the class next door] which were 134 cm, like, that, yeah, so, yeah. [Makes a face] [9 Nov, 1:42]

Elaine's suggestion prompted Sophie to rethink her plan and propose that they look beyond their two classes for what is the typical height more generally.

*Sophie:* Yes if, um, my idea is to say that if there was a typical height in grade 3, now let's just say, um, there are more people in our class that were 137 and more people in that class that were 136! Well, you should say my idea of typical is not what is now, not what's exactly, not what's [our data] *now*, I mean, not what has the *most*, but I'm saying what, what, um the *supposedly* measurement would be. Because not many people *regularly* are as tall as Charles, or Dan. [4:48]

Sophie's recognition that each class may have a different (single) most common height prompted her to go beyond a descriptive interpretation of typicality and consider an inferential interpretation. She proposed that a typical height for a year 3 student would not just be the most common for their two classes (“what is *now*”), but suggested a typical beyond this (“the *supposedly* measurement”) in the *population* (“people regularly”), contrasting these with heights of people who are unusually tall (outliers Charles and Dan).

The teacher took this as an opportunity to introduce the word *atypical* to students and, building on Sophie's proposal, asked the class to consider whether there are typical and atypical heights for year 3 students. This concept-laden vocabulary assisted students to adopt new language (“in the range” and “out of the range”) for talking about typical heights more generally as a normative range, even though this range was yet undefined.

After collecting height data from the class next door, students confirmed that the classes had several differences (minimum, maximum, most common, and frequencies of heights in the intervals 120, 130, 140 and 150 s cm). Patrice suggested that they combine the data from the two classes. In doing so, students were surprised to find that the most common height for the combined classes was 137 cm, different from the most common height in either class separately (132 and 128 cm). Combining the data initiated deeper discussions of typical height, with

students' proposals becoming increasingly more inferential. That is, the language students used in talking about their data began to refer to typical heights more broadly, beyond the data they had collected (Makar & Rubin 2009). From this point, students rarely returned to typical height as a single value and collectively worked under the assumption that the population estimate they were after was an interval.

In the final lesson, the teacher reminded students of the questions posed initially: *Is there a typical height in a year 3 class? If so, what could it be?* There had so far been implicit agreement that there was a typical height, but that they still needed to quantify an interval. Melanie tried grouping her data and noticed that in the middle range, frequencies were higher than for the outer intervals.

*Melanie:* I think I might be done. Well, I've just made a column of, it's just, like, all the heights. It's [combined with the other class] as well, and I've made tall, medium, and small columns, and most, and the most, and the column that's got the most amount of people is in the medium. [21 Nov, 6:24]

This larger middle group that emerged (cf. modal clump from Konold et al. 2002) prompted further discussion from the class. Sophie suggested that Melanie may have solved the problem without realising it.

*Sophie:* So did you say that medium had *the most*? [*Melanie:* Yes.] Well, I think that she's also arranged it in a certain way, but she hasn't figured it out yet. ... Medium must be the largest [group] ... [and that] shows what atypical and typical *is*, because the way she's done tall has, tall has only *a few* ... and I think tall might be "out of the range". I think the medium is actually "inside the range", because it has lots of columns, and it's the highest (untrans), and the first, 25 and over is [in] the range, 25 to 40 [125 to 140 cm]. [8:25]

Elaine worried, however, that if they collected data from a third or fourth class, the typical height may again be different. Melanie, however, argued that when they added the data for the second class, the middle group remained the largest, with the tall and small groups still containing only a few students. She hypothesised that this pattern of the normative height being in the middle would continue as they added more data: "if I put them all together, well, 130-something people [the middle group] keep getting bigger and bigger and bigger" [14:45] (see also Bakker, 2004 on growing samples).

#### 5.4 Typical height as representative around the world

The idea of an explicit typical range for heights concerned Emily, an Asian student who was significantly shorter than the other students. She explained that her mother, who was average height in her home country, always had to hem pants that she bought in Australia because here she was considerably shorter than the general population. Emily argued that "if you went around the world" [17:45], you'd find that typical heights would be different in different places. Elaine agreed.

*Elaine:* [In our class] the typical height is in the 30s, and in [the class next door], it's in the 20s. But ... I think if you go all around the world [to other] classes, I think there wouldn't be a typical height, 'cause there's different people, and they have different sizes. [22:05]

Elaine's response may reflect a reaction that is common when students initially wrestle with informal notions of sampling variability, feeling overwhelmed by the possibilities and retracting into a fairly relativistic position (Ben-Zvi, Aridor, Makar & Bakker 2012; Rubin, Bruce & Tenney 1990). Another possibility is that she was suggesting that there is not just one population for them to

consider, but multiple populations. The new possibility of no single typical height conflicted with the conclusion they had been building towards over the unit. This conflict prompted heated discussions when students returned to their collaborative groups to come up with a final conclusion for the inquiry question. They argued whether you can say there is a typical at all, as it may vary from country to country.

*Elaine:* If you go around the world, and you visit each year 3 class, there's people that are like Charles and taller (untrans)

*Charles:* That would take a long time! ...

*Elaine:* Some people are in the middle like us. Some people, like in Vietnam, they're really small. ... so there is actually no typical height. ...

*Charles:* I reckon there's only a typical height for every country. [small group discussion, 26:31]

Charles' response suggests that while there may be no single typical height overall, there could be a typical height that is representative of each country. This was an important insight to support his emerging concept of average as being a way of representing population groups, a notion that has been reported several times in the literature as being problematic for students learning statistics (e.g. Konold & Pollatsek 2002; Mokros & Russell 1995; Shaughnessy 2007) and one that had not been anticipated at this young age.

The class wrapped up the unit with a whole class discussion of whether there is a typical height. While they decided that there was no typical height for the world as a whole, they concluded the typical height for year 3 students in Australia to be in the 130s (130–139 cm). In doing so, they proposed a population estimate for the height of year 3 children in Australia based on the sample data they had collected. Their reasoning was extended by Barbara, who contended that this estimate relied on assumed probabilities.

*Barbara:* Since most people are in it, the more chances [are] that it is the typical height. [50:58]

Barbara's reasoning suggests her informal and preliminary understanding of concepts which underpin probability distributions, likelihood and the articulation of uncertainty when making population estimates from sample data. These are all key ideas for understanding statistical inference. Melanie cautioned, however, that it may change as “more people from different countries are coming into Australia” [53:30], recognising the dynamic character of the target population of their informal statistical inference.

## 6 Discussion

This study explored young children's investigation of the question, *Is there a typical height for year 3 children? If so, what is it?* where “typical” was used to engage children's familiarity with average. The results suggest a reconsideration of how notions of average can be developed by children exploring diverse conceptions of average in conjunction with informal inferential reasoning. This approach may support deeper conceptions of average than previously expected.

### 6.1 Average

The preliminary framework (Sect. 3) proposed concepts showing promise to deepen children's reasoning of average to incorporate multiple meanings, including reasonableness, outliers, most common (mode), group comparisons and inference. That framework is revisited here.

### 6.1.1 Reasonableness

Students' notion of average was conjectured to emerge initially out of everyday experiences with reasonableness (Mokros & Russell 1995). Engagement with the familiar context of heights allowed them to draw on tacit understandings of “sensible” heights. Their initial concept of average as reasonable was challenged and strengthened when they encountered heights that appeared unreasonable. These conflicting values, and their interest in explaining them, enabled the teacher to negotiate with them the need for accuracy when taking measurements. The concept of reasonableness therefore assisted in engaging students' meaning-making about average. This is important given the research documenting that students often ignore their own understanding of what makes sense in favour of more procedural approaches (Lavigne & Lajoie 2007; Mokros & Russell 1995).

In thinking about reasonable heights, the children's initial focus was on the heights of individual children. Konold and his colleagues (2004) described this as a case-value perspective of data where one attributes “average” as a feature of individual data. “To say that Sam is of ‘average’ height is to characterise this single case, and not necessarily the group the case is part of” (p. 30). Konold argued that a case-value perspective may explain why students object when an average is not equal to any points in their data. Therefore, while the concept of average as reasonable is a good start, the aim in statistics is to move students towards more aggregate reasoning. Had the class stopped at this point, there would probably been little reason for them to develop an understanding beyond this everyday case-value perspective of average.

### 6.1.2 Outliers

In recording height data from their class, students extended the concept of average from representing a reasonable height for an individual towards a meaning which could be defined as *not atypical*. The affordance of having a particularly tall child in the class evoked discussions of whether Charles' height was “in the range” or “out of the range” of typical heights. Konold et al. (2004) argued that consideration of outliers is an initial step towards looking at partial distributions as “an outlier involves locating it with respect to the other values in the distribution and thus involves a coordination of aggregate and individual perspectives” (p. 7).

### 6.1.3 Typical as most common

Modal values are frequently identified by students as the average of a dataset and a highly tenacious conception of average (e.g., Lavigne & Lajoie 2007; Mokros & Russell 1995). This concept of average emerged as students began collating the data, noting their frequencies. For some students, “most common” referred to a single value, such as when Barbara talked about three students at 132 cm being the most common height in the class. Other students identified typical values as a range. Whereas the mode as a range moves closer to a partial distribution, it still does not represent an aggregate view of data. The task that the teacher set may have promoted the idea of average as mode in the way that the questions were framed to suggest a single response: *Is there a typical height for a student in year 3? If so, what is it?* For a time, students did consider only the modal perspective of typical, for example Barbara (Sect. 5.2) using the single most frequent value as the typical height or when Sophie (Sect. 5.3) suggested taking a single student who was 132 cm next door to compare with students in the next class.

The students in this study likely maintained a sense of typicality as a mode when considering other meanings of typicality. This single experience would unlikely be sufficient to shake the tenacity of mode as students' primary conception of average.

#### 6.1.4 Comparing groups

When asked whether the typical heights in their class would also be typical in the class next door, shifts occurred in students' reasoning about typical. By comparing their data to the class next door, they encountered informal notions of sampling variability when the shortest, tallest and most common values did not align with their own class data. This experience, met earlier in the year in the *Hand Span* unit, provoked them to make sense of these differences. One aim of statistics is for students to think about average as representative of a dataset when comparing groups, so there is some evidence that the initial experience of managing the variability between samples may have allowed students to begin to consider this perspective.

#### 6.1.5 Informal statistical inference

Sophie's revelation of average as a “*supposedly* measurement” that was more indicative of “people regularly” was a critical moment in the class. Patrice's suggestion to combine the two classes may have been her attempt to get a sense of an aggregated typical, aiming for Sophie's “*supposedly* measurement”. The modal value in the combined classes differed from the modes of the classes individually (132 and 128 cm); these conflicts between students' expectations and the messages in the data were critical points of learning as they probably triggered students' motivation to dig further (Dewey 1938; Watson 2007). In discussing the combined data with “the most amount of people” in the medium column, Melanie and Sophie provided further evidence that they were starting to think about the data as an aggregate. Melanie argued that the middle group would “keep getting bigger and bigger and bigger” as they added more data, suggesting her expectation that the distribution categories would stabilise (cf. the Law of Large Numbers).

By the end of the unit, students were beginning to envision typical height as representative of the population. When Elaine argued that the typical height in Australia was not representative of the typical height in other parts of the world, this shook students' confidence in any existence of a typical height. Charles' suggestion that “there's only a typical height for every country” appeared to settle students back into finding the typical height for their own country. The students appeared to be working inferentially in this final stage of the unit as they were working towards a notion of typical height in greater population (beyond their class), using the data they collected as evidence for their arguments and articulating uncertainty about what might be the typical height (Makar & Rubin 2009).

#### 6.1.6 Summary and caveats

The students' understanding of average became more mathematically precise through the unit as their understanding of typical progressed from initial descriptions of their own data to inferences about populations beyond their classroom. This paper suggests that the framework used may show promise as the basis of a learning trajectory for primary students; however, this approach would require additional research in diverse settings to be able to make this claim more firmly. The results presented here are neither intended to suggest that students formalised these conceptions of average nor that they left behind each notion as they moved onto the next

one. While these findings suggest that students encountered multiple meanings of average and typicality, development of these concepts requires that students have experiences which they can tie to their growing understandings as they are revisited multiple times (Harel & Koichu 2010; McGowen & Tall 2010). Finally, the aim in this study was not to move students towards thinking about the arithmetic average, but rather to experience informally these multiple conceptions of average. The literature argues that the problem has been that the arithmetic mean has been given too much attention before richer notions of average have been developed. There is no claim there that students have developed firm understandings of any of the above notions of average nor that they have (or even should) let go of the notion of mode. Instead, the paper claims they have experienced a number of meanings for average that may provide them with a richer foundation for exploring these conceptions in more depth and formally in later years.

## 6.2 Informal inferential reasoning

Informal inference has been an emerging concept in statistics education for several years and could be argued as the “new statistics” for building students' statistical reasoning across the years of schooling. The results of this study suggest that elements identified to support informal inferential reasoning (Makar et al. 2011) were both present and potentially critical for students' understanding of average and reasoning inferentially about the problem they were addressing. Each of these elements is discussed below.

### 6.2.1 *Inquiry-based environment*

The classroom from which the data were drawn is one which developed mathematical inquiry. That is, the question that students were addressing was ill-structured and ambiguous. This ambiguity required them to negotiate and mathematise their ideas as they worked towards a solution. Norms of collaboration and public debate are central to an inquiry-based environment (Cobb 1999). Students had been developing these norms through the year and this could be observed in the way that they critiqued and probed each other's ideas, built their ideas on other's and created new ways of talking about their emerging understandings of heights that were “sensible”, “in the range” and “typical around the world”. Within this inquiry-based environment, students were encouraged to generate, reconsider and build concepts through peer interaction as they thought aloud, modified ideas, argued and shaped their thinking about statistics.

### 6.2.2 *Statistical concepts and tools*

The children utilised statistical concepts not normally introduced at this age, suggesting that young children may be able to grasp informal notions of outliers, group comparisons, sampling variability, representativeness, populations and informal inference within an inquiry-based environment. Although they did not consolidate these statistical concepts, their informal use of these concepts suggests that they were building foundational knowledge that could enable further development when they encountered them again in more complex problems. Without these statistical concepts, they probably would not have been able to reach the depth of understanding about the contextual problem which they were addressing. In this way, the statistical concepts were critical to their developing inferential reasoning about the typical heights of children their age.

### 6.2.3 Conflict

Over the course of the unit, students encountered multiple conflicts between their expectations and the data. Initially, their sense of reasonableness emerged from data points that did not seem to fit values that they expected to find. Charles' unusual height assisted them in comparing their notion of typical with the atypical heights that did not fit this expectation; out of this experience, their language of “in the range” and “out of the range” assisted them to develop meaning about heights that they tacitly categorised as typical. The anticipation and confirmation that the typical heights in the neighbouring class did not match their own provided impetus to further consider what typical could mean more broadly. Finally, Emily's concern that their emerging definition of typical height did not match her understanding of typical within her cultural background sent students into vigorous discussion of whether there was a single notion of typical around the world. Discussing heights of children was not an abstract notion, nor was it just about describing data. The familiarity of having Charles (as an outlier), Elaine (as representing the typical height in their own class) and Emily (as representing the existence of wider populations) in the class generated discussion about specific conceptual tools that emerged to explain and incorporate these notions. In each of these cases, conflicts that students encountered between their expectations and the messages interpreted in the data provided them with opportunities to debate, clarify and seek to resolve their ideas. These conflicts arose from the complexity of the authentic situation and a culture of inquiry which encouraged statistical concepts to emerge through debate and deliberation. A key aspect here was the teachers' skill in provoking students' reasoning and developing a class culture which valued substantive conversation.

The evidence in this study suggests that these three elements were critical to supporting these students' informal inferential reasoning, and further, were key elements that enabled them to reach the level of depth of understanding of average.

## 7 Conclusion

This paper operationalises new research in statistics education by illustrating how young children can use informal inferential reasoning to develop rich conceptions of average. The recent move towards using informal inferential reasoning as a unifying theme in statistics education provokes a more holistic approach to teaching and learning in statistics education (Bakker & Derry 2011) to hopefully combat the “meanmedianmode” syndrome (Bakker 2004). These new research directions in statistics education align with broader research trajectories which embrace complex, multidimensional process-oriented research (Harel & Koichu 2010; Mercer 2010). This paper describes only a few lessons in a single classroom and is not intended to make broad generalisations about the level of all students' learning in the class. Neither does it address the difficult task of how to support teachers in creating an inquiry-based classroom (Makar 2011) nor in helping teachers to support students' informal inferential reasoning (Pfannkuch 2005). However, it provides insight into possibilities for developing children's statistical concepts and reasoning when young children are allowed to grapple with ill-structured problems. Using this classroom as an illustration, this study can help teachers begin to undertake and design messy problems to get beyond uncritical conceptions of average and suggest a different pedagogy to teaching statistics.

**Acknowledgments** This study was funded by the Australian Research Council (LP0776703), Education Queensland and The University of Queensland. An abbreviated version of these data were presented at the 2009 annual conference of the *Mathematics Education Research Group of Australasia*. The author thanks Debra



McPhee for her insightful teaching and collaboration on the MERGA version of this paper. Note that the children whose work appears in this paper own their IP.

## References

- Ainley, J., Pratt, D., & Nardi, E. (2001). Normalising: Children's activity to construct meanings for trend. *Educational Studies in Mathematics*, *45*, 131–146.
- Allmond, S., Wells, J., & Makar, K. (2010). *Thinking through mathematics: Engaging students with inquiry-based learning (Books 1-3)*. Melbourne: Curriculum Press.
- Bakker, A. (2004). Reasoning about shape as a pattern in variability. *Statistics Education Research Journal*, *3*, 64–83.
- Bakker, A., & Derry, J. (2011). Lessons from inferentialism for statistics education. *Mathematical Thinking and Learning*, *13*, 5–26.
- Ben-Zvi, D., Aridor, K., Makar, K., & Bakker, A. (2012). Students' emergent articulations of uncertainty while making informal statistical inferences. *ZDM—The International Journal on Mathematics Education (Zentralblatt für Didaktik der Mathematik)*, *44*, 913–925.
- Ben-Zvi, D., Gil, E., & Apel, N. (2007, August). *What is hidden beyond the data? Young students reason and argue about some wider universe*. Paper presented at the Fifth International Forum for Research on Statistical Reasoning, Thinking and Literacy, University of Warwick, UK.
- Cobb, P. (1999). Individual and collective mathematical development: The case of statistical data analysis. *Mathematical Thinking and Learning*, *1*, 5–43.
- Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*, 9–13.
- Dewey, J. (1938). *Logic: The theory of inquiry*. New York: Henry Holt & Company.
- Dierdorff, 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.
- Flick, U. (2009). *Introduction to qualitative research* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Friel, S. N., O'Connor, W., & Mamer, J. D. (2006). More than “Meanmedianmode” and a bar graph: What's needed to have a statistical conversation. *Thinking and reasoning with data and chance: Sixty-eighth NCTM yearbook* (pp.117-137). Reston, VA: NCTM.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, *70*, 1–25.
- Ginat, D., & Wolfson, M. (2002). *On limited views of the mean as point of balance*. Paper presented at the 26th annual conference of the International Group for the Psychology of Mathematics Education. Norwich, UK: University of East Anglia.
- Gould, R. (2011). *Informal reasoning for informal samples: Statistics in the modern world*. Paper presented at the Seventh International Forum for Research on Statistical Reasoning, Thinking and Literacy (SRTL7). Texel, The Netherlands: SRTL.
- Harel, G., & Koichu, B. (2010). An operational definition of learning. *Journal of Mathematical Behavior*, *29*, 115–124.
- Konold, C., Higgins, T., Russell, S.J., & Khalil, K. (2004). *Data Seen through Different Lenses*. Unpublished manuscript. <http://srri.umass.edu/files/konold-2004dst.pdf>, Accessed 9 March 2009.
- Konold, C., & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, *33*, 259–289.
- Konold, C., Robinson, A., Khalil, K., Pollatsek, A., Well, A., Wing, R., et al. (2002). Students' use of modal clumps to summarize data. In B. Phillips (Eds.), *Developing a statistically literate society. Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg: International Association for Statistical Education. The paper is available at [http://iase-web.org/documents/papers/icots6/8b2\\_kono.pdf](http://iase-web.org/documents/papers/icots6/8b2_kono.pdf)
- Lampen, C. E. (2013). *Learning to teach statistics meaningfully*. Unpublished doctoral dissertation, University of Witwatersrand, South Africa.
- Lavigne, N., & Lajoie, S. (2007). Statistical reasoning of middle school children engaged in survey inquiry. *Contemporary Educational Psychology*, *32*, 630–666.
- Makar, K. (2011). *Learning over time: Pedagogical change in teaching mathematical inquiry*. In Julie Clark, Barry Kissane, Judith Mousley, Toby Spencer & Steve Thornton (Eds.), *Mathematics: Traditions and [New] Practices* (pp. 25-37) (Proceedings of the 34th annual conference of the Mathematics Education Research Group of Australasia and the Australian Association of Mathematics Teachers). Adelaide: AAMT and MERGA. [http://www.merga.net.au/publications/counter.php?pub=pub\\_conf&id=1642](http://www.merga.net.au/publications/counter.php?pub=pub_conf&id=1642)

- Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning*, 13, 152–173.
- Makar, K., & Ben-Zvi, D. (2011). Editorial: The role of context in developing reasoning about informal statistical inference. *Mathematical Thinking and Learning*, 13, 1–4.
- Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In C. Batanero et al. (Eds.), *Teaching statistics in school mathematics. Challenges for teaching and teacher education* (pp. 347–358). New York: Springer.
- Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, 8, 82–105.
- McGowen, M. A., & Tall, D. O. (2010). Metaphor or met-before? The effects of previous experience on practice and theory of learning mathematics. *Journal of Mathematical Behavior*, 29, 169–179.
- Mercer, N. (2010). The analysis of classroom talk: Methods and methodologies. *British Journal of Educational Psychology*, 80, 1–14.
- Mokros, J., & Russell, S. J. (1995). Children's concept of average and representativeness. *Journal for Research in Mathematics Education*, 26, 20–39.
- Papariostodemou, E., & Meletiου-Mavrotheris, M. (2008). Developing young students' informal inference skills in data analysis. *Statistics Education Research Journal*, 7, 83–106.
- Pfannkuch, M. (2005). Probability and statistical inference: How can teachers enable learners to make the connection? In G. Jones (Ed.), *Exploring probability in school: Challenges for teaching and learning* (pp. 267–294). Dordrecht: Kluwer Academic Publishers.
- Pfannkuch, M. (2006). Informal inferential reasoning. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg: International Statistical Institute. [http://iaseweb.org/documents/papers/icots7/6A2\\_PFAN.pdf](http://iaseweb.org/documents/papers/icots7/6A2_PFAN.pdf)
- Pratt, D., & Ainley, J. (2008). Introducing the special issue on informal inferential reasoning. *Statistics Education Research Journal*, 7, 3–4.
- Reitman, W. (1965). *Cognition and thought: An information-processing approach*. New York: Wiley.
- Roth, W.-M. (1995). From “wiggly structures” to “unshaky towers”: Problem framing, solution finding, and negotiation of courses of actions during a civil engineering unit for elementary students. *Research in Science Education*, 25, 365–381.
- Rubin, A. (2005). Math that matters: The case for replacing the algebra/calculus track with data literacy—a critical skill for modern life. *Threshold Magazine: Exploring the Threshold of Education*, Spring 2005 issue, 22–25, 31.
- Rubin, A., Bruce, B., & Tenney, Y. (1990, April). *Learning about sampling: Trouble at the core of statistics*. Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.
- Rubin, A., Hammerman, J., & Konold, C. (2006). Exploring informal inference with interactive visualisation software. In A. Rossman & B. Chance (Eds.), *Working Cooperatively in Statistics Education. Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg: International Statistical Institute. [http://iaseweb.org/documents/papers/icots7/2D3\\_RUBI.pdf](http://iaseweb.org/documents/papers/icots7/2D3_RUBI.pdf)
- Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 97–115). Cambridge: Cambridge University Press.
- Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol 1, pp. 957–1009). Greenwich, CT: Information Age Publishing and NCTM.
- Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Mahwah NJ: Lawrence Erlbaum.
- Watson, J. (2007). The role of cognitive conflict in developing students' understanding of average. *Educational Studies in Mathematics*, 65, 21–47.
- Watson, J., & Moritz, J. (1999). The beginning of statistical inference: Comparing two datasets. *Educational Studies in Mathematics*, 37, 145–168.
- Wild, C., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67, 223–265.
- Wolfe, C. R., Reyna, V. F., & Brainerd, C. J. (2005). Fuzzy-trace theory: Implications for transfer in teaching and learning. In J. P. Mestre (Ed.), *Transfer of learning from a modern multidisciplinary perspective* (pp. 53–87). Greenwich, CT: Information Age Publishing.