Average Revisited in Context

Jane Watson
University of Tasmania
<Jane.Watson@utas.edu.au>

Helen Chick
University of Tasmania
<Helen.Chick@utas.edu.au>

This paper analyses the responses of 247 middle school students to items requiring the concept of average in three different contexts: a city's weather reported in maximum daily temperature, the number of children in a family, and the price of houses. The mixed but overall disappointing performance on the six items in the three contexts indicates the need for concerted efforts to link numeracy across the curriculum as required in the new *Australian Curriculum*.

The concept of average, reflected mainly in relation to the arithmetic mean, was one of the first areas of mathematics education research that touched on statistics. Starting in the 1980s, researchers considered tertiary students' understanding of the weighted mean, with some success reported for interventions to improve performance (e.g., Mevarech, 1983). Somewhat later, other researchers considered school students' appreciation of properties of the arithmetic mean (e.g., Strauss & Bichler, 1988), culminating in the seminal work of Mokros and Russell (1995) in considering average more generally in terms of ideas related to mean, median, and mode but categorised as representative or non-representative. At the same time, procedural issues related to the algorithm for calculating the mean and the concept of the mean as a balance point were being considered by others (e.g., Cai, 1998). The first longitudinal studies of school students' understanding of average employing a cognitive development model were reported by Watson and Moritz (1999, 2000). After 2000 the interest moved to teachers, with Jacobbe and Carvalho (2011) providing an overview of teachers' and pre-service teachers' understanding.

The word "average," *per se*, is not used in the *Australian Curriculum: Mathematics* (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2011) in any descriptors or elaborations associated with Data Representation. The word "mean" appears at Years 7, 8, 9, and 10A; "median" appears at Years 7, 8, and 9; and "mode" appears at Year 7. The most meaningful descriptors and elaborations are presented for Year 7 (see Table 1). Centre, spread, and outliers are important statistical words in the elaborations, as are the references to some sort of context, in this case "compar[ing] land use in the local municipality" and "connecting them to real life." In Year 8 there is reference to "using mean height for a class," whereas in Year 9 the elaboration "comparing the rainfall in various parts of Australia, Pakistan, New Guinea and Malaysia" could employ the mean and median, mentioned elsewhere in Year 9.

The occasional insertion of context within Mathematics relates to the seven General Capabilities required across the entire Australian curriculum. One of the seven is Numeracy, and each other curriculum area—for example, History or Science—includes a statement on how the Numeracy Capability is developed within its discipline. Across all disciplines students

> need to recognise the mathematical basis of authentic problems and engage constructively in their solution. The identification of mathematical demands in learning areas enables students to:
>
> - transfer their mathematical knowledge and skills to problem solving in those learning areas
> - recognise the interconnected nature of mathematical knowledge, learning areas and the wider world
> - become confident and willing users of mathematics in their lives.

In J. Dindyal, L. P. Cheng & S. F. Ng (Eds.), Mathematics education: Expanding horizons (Proceedings of the 35th annual conference of the Mathematics Education Research Group of Australasia). Singapore: MERGA.

Table 1
Year 7 Mathematics Curriculum extracts (ACARA, 2011)

Data representation and interpretation	Elaborations
Calculate mean, median, mode and range for sets of data. Interpret these statistics in the context of data (ACMSP171)	• Understanding that summarising data by calculating measures of centre and spread can help make sense of the data • Calculating mean areas set aside for parkland, manufacturing, retail and residential dwellings to compare land use in the local municipality
Describe and interpret data displays and the relationship between the median and mean (ACMSP172)	• Using mean and median to compare data sets and explaining how outliers may affect the comparison • Locating mean, median and range on graphs and connecting them to real life

With context playing an explicit role in the *Australian Curriculum* and in light of previous research the following research question is considered in this report.

> What levels of understanding do middle years students display in relation to linking their definitional knowledge of average to the context within which it is placed?

Methodology

Instruments

The items about "average" used in this study were adapted from earlier research. They are presented in Figure 1 in the order in which they appeared on a larger survey. The first item, Q5, about the average family having 2.3 children, was originally used as part of an interview protocol by Watson and Moritz (2000). In the original study this question followed a question about what it means for a family to have 2.3 children, which was helpful in setting the context for the item as used there. The second and third survey items, Q9 and Q10, were part of a protocol used by Watson and Kelly (2005) exploring students' understanding of variation in the context of a city's daily maximum temperature over a year. As discussed earlier, spread, outliers, and range provide context for considering the mean and its contribution to an overall description of a distribution. The last three items, Q21 to Q23, were first used by Watson and Moritz (1999) in a survey format with 1654 students in grades 6, 8, 9, and 11. The context of median house prices is one of the most common applications of average in social settings, especially as a contrast to mean. Due to time constraints not all study participants completed the final three items.

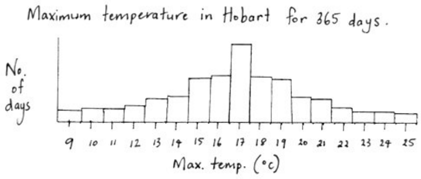
Sample

The 247 students who completed the survey were part of the StatSmart project (Callingham & Watson, 2008; Callingham, 2010) and had completed one similar survey previously during the project, including Q21 to Q23 but not Q5, Q9, or Q10. The students came from three Australian states (Tasmania, 136; Victoria, 64; South Australia, 47). The number of students at each year level is given in Table 2. For the purpose of analysis, the year groups were combined in pairs, 6/7, 8/9, and 10/11. In one state Year 7 was a primary grade and there were relatively few students in Year 6. Years 8 and 9 are, according to the curriculum, appropriate for consolidating the concept of average. There were very few Year 11 students in the study. Overall there were 47% females and 53% males.

Item 5: The average number of children in 10 families in the neighbourhood is 2.3. One family with 5 children leaves the neighbourhood. What is the average number of children per family now? Show your work here.

Item 9: Some students watched the news every night for a year, and recorded the daily maximum temperature in Hobart. They found that the average maximum temperature in Hobart was 17° C

Item 10: Ben drew the following graph to show the maximum temperatures in Hobart throughout the year. Explain what the graph shows.



This article appeared in a newspaper.

Hobart defies homes trend

AGAINST a national trend, Hobart's median house price rose to $88,200 in the March quarter - but, Australia-wide, the average wage-earner finally can afford to buy the average home after almost two years of mortgage pain.

Item 21: What does "average" mean in this article?
Item 22: What does "median" mean in this article?
Item 23: Why would the median have been used?

Figure 1. Items used in survey.

Table 2
Students in Each Year

Year	6	7	8	9	10	11	All
N	14	56	36	59	78	4	247

Analysis

The rubrics used to score responses to the six items are given in Table 3. The codes reflect the influence of the SOLO Model in recognising greater structural complexity at higher levels but also the required correctness or appropriateness of responses. The rubrics for Q5, Q9, and Q10 are based on those used in earlier interview settings by Watson and Moritz (2000) and Watson and Kelly (2005). Those for Q21 to Q23 are found in Watson and Callingham (2003).

Results

A summary of the percent of responses by item and level is given in Table 4. For Q21 to Q23 the percentages refer to the number of students reaching these items. NA records the number of missing students. Performance across the codes for the items are similar, but with the mean performance of Year 8/9 students better than Year 6/7 students on all items and marginally better than the Year 10/11 students on all but Q9. Because of the difference in pattern of performance across the years for Q9 and Q10 and the closer relationship of Q5 and Q21 to Q23 to the definitions of mean and median in the *Australian Curriculum* (ACARA, 2011) these four items are presented first, followed by Q9 and Q10.

Table 3

Rubrics for the Six Survey Items

Code	Criteria
Question 5, 2.3 children	
0	No response; incorrect answer with no explanation or unintelligible reasoning
1	Partial attempt – can recognize some aspect of the problem
2	Correct answer with no explanation or plan without explicit answer
3	Correct answer with appropriate explanation
Question 9, 17 degrees	
0	No response; imaginative or idiosyncratic comments
1	A comment about an aspect of temperature OR a statement describing 17 on a continuum
2	A comparison of temperature with other places
3	A comment about the temperature with an acknowledgment of variation focusing around 17; or an explicit reference to variation away from 17
Question 10, 17 degrees and graph	
0	No response; idiosyncratic responses or misreading details of the graph
1	A comment about the shape of graph; however, no understanding of graph purpose evident
2	Statement about frequency and understands graph purpose, but no focus on importance of 17
3	Focus on 17 only: more frequent temperature
4	Statement about graph's frequency and purpose, and reference to the importance of 17
Question 21, "average" in article	
0	No response; No idea of central tendency, often tautological
1	Single idea not related to context
2	Describes central tendency for a data set or method of obtaining average from a data set (maybe related to context) – (needs to state both central tendency and the associated data set)
Question 22, "median" in article	
0	No response; No idea of central tendency, often tautological
1	Single idea not related to context
2	Describes the central tendency for a data set or the method of obtaining the median from a data set (sometimes related to context)
Question 23, Why "median"	
0	No response; Response that does not refer to question (e.g., language/price)
1	Usefulness or fairness (without explicit mention of outliers)
2	Mention of outliers or extreme values

Table 4

Responses at Each Code Level and Mean for Items and Year.

Year Item and code levels	6/7 n = 70	8/9 n = 95	10/11 n = 82
Q5, 2.3 children (0, 1, 2, 3)	(73, 17, 4, 6) mean = 0.40	(34, 17, 6, 43) mean = 1.57	(45, 20, 4, 32) mean = 1.20
Q9, 17 degrees (0, 1, 2, 3)	(59, 23, 6, 13) mean = 0.73	(43, 37, 4, 16) mean = 0.93	(56, 17, 4, 23) mean = 0.94
Q10, 17 degrees and graph (0, 1, 2, 3, 4)	(30, 11, 19, 26, 14) mean = 1.81	(15, 13, 33, 17, 23) mean = 2.21	(26, 9, 21, 20, 26) mean = 2.11
Q21 "average" in article (0, 1, 2, [NA*])	(55, 34, 11, [25]) mean = 0.57	(16, 61, 23, [21]) mean = 1.07	(31, 51, 18, [21]) mean = 0.87
Q22 "median" in article (0, 1, 2, [NA*])	(66, 16, 18, [26]) mean = 0.52	(31, 34, 35, [21]) mean = 1.04	(28, 47, 25, [25]) mean = 0.96
Q23 why "median" (0, 1, 2, [NA*])	(86, 14, 0, [26]) mean = 0.17	(62, 24, 14, [21]) mean = 0.51	(73, 18, 9, [26]) mean = 0.36

(x_0, x_1, x_2, …, x_n) indicates the percent attempting the item who performed at levels (0, 1, 2, …, n)

* Number who did not reach the item.

Item Q5, 2.3 Children

Of the half of the 247 responses that were coded 0, most (89) put question marks, said they did not know, or left the space blank. Of the 37 responses coded 0 that wrote something, 13 put down a number different from 2 with no explanation (e.g., 1.3, 1.6, 2.1, 5, 1.15). Two responses were drawings of boxes with tallies (likely to represent families with children) with no conclusion. Two responses questioned the existence of 0.3 of a child. Ten responses provided calculations that could not be interpreted (e.g., 2.3 ÷ 9 = 4.23, 10 ÷ 4 = 2.3). Three responses provided a written answer indicating that the average did not change because "the maximum number of children would be 5 to start with most likely," because "the family had more than the average amount of children," or "2.3 cause it doesn't say that the family was the only one that had five kids."

Of the 47 responses coded 1, 24 worked with the number 23 (10 x 2.3), subtracted 5, but divided by 10 rather than 9, obtaining an answer of 1.8. Other answers had errors in calculations using numbers that indicated the students knew or selected numbers relevant to the solution, perhaps not completing the process. Six of these answers were larger than 2.3, indicating that students did not check their answers, as the question implied a reduction in the average due to more than the average being removed. Four responses showed some understanding of the meaning of average but claimed that there was not enough information provided to work out the answer (e.g., "It would depend on how many children the other 9 houses had").

Of the responses coded 2, three responses showed more complete intuition but did not actually calculate the new average (e.g., "Less than 2.3," "Decreases, because a high number was taken out which was keeping the average high"). Nine others gave an answer of 2.0 but with no explanation or unclear justification. Of the 71 students who showed appropriate calculations to reach the answer of 2.0 (coded 3), there was a variation in the amount of detail provided. Eight of these responses created 10 families with appropriate numbers of children to equal 23, with one of 5; that family was removed and the average shown for the 9 remaining. It would appear that the context supported these students in solving the problem. Others followed the procedure suggested above dividing by 9.

Items 21 to 23, Average and Median

As shown in Table 4 these items were more difficult for students than the others. Across Q21 to Q23, of the responses coded 0, an increasing number wrote that they did not know or could not remember the concept involved or why the median was relevant (20 for Q21, 48 for Q22, and 66 for Q23). A further 12 responses to Q21 were tautological, as was 1 response to Q22. Other answers to Q21 that were coded 0 focussed on irrelevant aspects: "average means decent," "average means higher number," "a house that could be any sort of house." Q22, about the meaning of median, had responses such as the following: "the median means the price of houses rose," "most popular," "maybe the main price." The modal idea occurred on several occasions. Many of the responses to Q23 did not focus on dealing with outliers when explaining why the median was used: "because people are most likely to see a house that costs 88 700," "to determine if the average worker can afford the house," "because the average is already used," "to show an approximate house price in Hobart," "to add in statistical stuff and make it interesting."

For Q21, on the meaning of average, responses did not have to focus on a particular meaning of average. Responses coded 1 reflected a single appropriate aspect of any type of average: "the normal wage earner," "the most common wage earner," "in the middle." A code of 2 was given for responses that added to this the context and/or discussion of the data set from which the average is obtained: “The average means the workers whom are not high nor low, but middle,” “It’s all the data added and divided by the number of data,” “The average house means not an expensive house but not a cheap one either. In other words the common home.”

Similarly, for Q22 on the meaning of median, code 1 responses reflected a single aspect of the concept: “I think it means the same as average,” “standard house price.” Code 2 responses were more specific about the properties of median, although perhaps more colloquial than a text book definition: “Median is the amount of money used for the houses that sold in between the highest and lowest,” “Median means the middle house price of the March quarter,” “The complete middle of ranked data.”

For Q23, code 1 responses reflected the usefulness of the median without referring to outliers: “Because the median is the middle range amount that people spend,” “So that readers don’t think the house prices are too high or too low,” “Average would be inaccurate.” Code 2 responses acknowledged the need to avoid outliers: “To exclude any extremely high or low prices,” “Because if the mean was used, the price of a really expensive house would put up the mean. The median is more accurate.”

Items Q9 and Q10, Weather

Items Q9 and Q10 were more open-ended than the other items, inviting students to reflect on the contextual meaning of an average rather than its mathematical definition. Fewer students than for Q5 left these items blank or said they did not know (30 for Q9 and 18 for Q10), suggesting that the context and question encouraged engagement. For many of the items within year groups the mean performance compared to the total possible score was better for Q10.

For Q9—which asked about average temperature—there were 13 tautological responses. Other responses coded 0 for Q9 presented views about the weather not related directly to the average maximum temperature or involving a misinterpretation of the average. These included: “This tells us that the temperature rarely rises over 17^{o}C,” “That the temperature is usually 17^{o}C maximum,” “We might have had a drought.”

Code 1 responses to Q9 contained a single comment about an aspect of temperature or about 17^{o} on a continuum: “It tells us that most days the temperature in Hobart is 17^{o}C or close to it,” “That it is 17^{o}C. We have had some cold days, we’ve had some hot days.” Code 2 recognised responses that compared Hobart temperatures with other places. As seen in Table 4, there were few of these: “Hobart’s weather for maximum temperature is around 17^{o}C, not going up much like the mainland,” “Well when you think of WA their average would most likely be in the 20 range. So that tells us that a lot of the time it is unpredictable but I would say Hobart is a cold place.”

In contrast, the highest level of response (Code 3) acknowledged variation close to or away from 17^{o}C: “the average maximum temperature is not very warm. It’s probably because Hobart has cold winter days and not-so-hot summer days,” “It may be hot in summer but it's colder for the rest of the year which brings the whole average down,” “It is quite cool. Probably ranging from about 0^{o} → 28^{o} mostly in the 17^{o} area,” “Even though there would be plenty of hot ones there were also cold ones to keep the average down.”

Item Q10, which referred to a graph of the year’s daily maximum temperatures, was the most open-ended of the six items analysed here. Its purpose was to assess the link between the knowledge that the average maximum for the year was 17^{o}C and a graph depicting the frequency of temperatures throughout the year. The graph was created as a "student response" and not based on actual data. Student responses coded 0 were idiosyncratic or misread details of the graph: "It's hot around 16 17 18," "17^{o}C was the highest," "The graph shows that the maximum degrees in Hobart is 17^{o}C and it shows that it is a vertical graph," "That it's 'averaged' by mode."

Responses coded 1 noted the shape of the graph but misinterpreted its purpose: "The temperature rises then drops, " "It shows the mean, mode and median of Hobart's weather throughout the year," "The most frequent temperature," "In the middle of the year the temp has gotten hotter and in the beginning and end it dropped," "It goes in a pattern, starts low, goes up, then low again." Code 2 responses made statements about frequency and the purpose of the graph but not the importance of 17^{o}C: "How high the temp got to in ^{o}C and how many days it got to that temp," "That the temperature was hardly ever extremely high or extremely low, it stayed in the middle," "It shows the range of temperature maximums: 9-25. It also shows how many days were a certain temp relative to other temps," "That we are mainly average temperature but we do have a few hot days and few cold days."

Code 3 responses focussed on 17^{o}C only as the more frequent temperature, sometimes with additional erroneous interpretation: "That it was 17^{o} a lot of days throughout the year," "It is most likely to be 17^{o} in Hobart," "This graph shows that 17^{o}C was a very popular temperature in Hobart. That the temperature goes up to 17^{o}C then down," "I think it shows that the average maximum temperature was 17^{o}C that year." Code 4 responses, in contrast, attended to multiple aspects of the graph, and made statements about frequency and the purpose of the graph including acknowledgment of the importance of 17^{o}C. Examples of these responses include: "It rarely gets to 9 and 25^{o}C but it does a lot get to 17^{o}C," "It shows us the maximum temperature did not exceed 25^{o} or come below 9^{o}. It was 17 almost twice as often as it was 15, etc, though it doesn't actually display the no. of days," "It show that most of the time the temperature is around 15-19^{o}C, with 17^{o}C being the most. This is most likely why the average temperature is 17^{o}C as that is the range where the temperatures mostly were," "That 17 was the most common temp but it varied alot," "That a majority of the days were 17^{o} and many were close to 17^{o}," "That most days were seventeen. We have less days in each temperature as we move away from 17^{o}C."

Discussion

The outcomes of the analysis of these three contexts point to some intriguing conclusions and messages for the curriculum, assessment, and teachers. When students are asked specific procedural or definitional questions in context (Q5, Q21-Q23) they do not perform as well as when they are asked more general questions placing central tendency in a context that encourages the engagement with variability in the overall context (Q9, Q10). Given the description and elaboration in Table 1 from the curriculum, students at least from Year 8 would be expected to perform better on all items, and certainly on Q5, Q21-Q23 in particular, than seen in this study. These four items reflect an emphasis—on real world properties of mean and median—that has been in the curriculum for many years. Q9 and Q10 reflect a more recent focus on the contribution of variation. The fact that the students in this study have been more engaged and successful with Q10 would appear to support Watson's (2005) view that although expectation and variation are the foundations of the Chance and Data curriculum, variation is more intuitive, developing earlier.

At least two implications arise for teaching. On no items is there a noticeable continued improvement in performance from Years 8/9 to 10/11. Further explicit engagement with the concepts appears to be required. If teachers can appreciate the diversity in the levels of comprehension found here, and the aspects of understanding that are more difficult for students to reach, then they should be able to devise remedial strategies to assist students to come to a stronger understanding. The limited nature of some responses suggests that greater care and attention should be given to encouraging in-class discussion of how statistical information complements and enhances understanding of real world contexts, and to sharing different appropriate strategies for making sense of data in context.

The writers of assessment items, for example for NAPLAN, may need to rethink what questions are asked and how, if they want to encourage students to have a wider view of the application of the concept of average than just the definitions and procedures. It is easy to write an item like Q5 but not as easy to devise good multiple choice distracters for items like Q9 and Q10. In the future the demands of the *Australian Curriculum* will require both procedural and conceptual understanding, and due consideration of context.

Teachers who are aware of the General Capabilities in the *Australian Curriculum* also need to be aware of the need to satisfy not only the requirements of Mathematics but also the more general Numeracy needs of other areas of the curriculum and the goal to "become confident and willing users of mathematics in their lives" (ACARA, 2011). Tasks related to Q9 and Q10 are more likely to be meaningful for these aims than items like Q5.

Acknowledgements

This research was partially funded by ARC Linkage Grant No. LP0669106.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2011). *The Australian Curriculum: Mathematics, Version 1.2, 8 March 2011.* Sydney, NSW: Author.

Cai, J. (1998). Exploring students' conceptual understanding of the averaging algorithm. *School Science and Mathematics, 98*, 93-98.

Callingham, R., & Watson J. M. (2008). Overcoming research design issues using Rasch measurement: The StatSmart project. In P. Jeffery (Ed.), *Proceedings of the AARE annual conference, Fremantle, December, 2007*. Available at http://www.aare.edu.au/07pap/cal07042.pdf

Callingham, R. (2010). Trajectories of learning in middle years' students' statistical development. Refereed paper in C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society*. (Proceedings of the 8th International Conference on the Teaching of Statistics, Ljubljana, Slovenia, July). [CD-ROM] Voorburg, The Netherlands: International Statistical Institute.

Jacobbe, T., & Fernandes de Carvalho, C. (2011). Teachers understanding of average. In C. Batanero, G. Burrill, C. Reading, & A. Rossman (Eds.), *Teaching Statistics in School Mathematics – Challenges for Teaching and Teacher Education* (pp. 199-209). New York: Springer.

Mevarech, Z. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics, 14*, 415-429.

Mokros, J., & Russell, S. J. (1995). Children's concepts of average and representativeness. *Journal for Research in Mathematics Education*, *26*, 20-39.

Strauss, S., & Bichler, E. (1988). The development of children's concept of the arithmetic average. *Journal for Research in Mathematics Education*, *19*, 64-80.

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

Watson, J. M., & Callingham, R. A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3-46.

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

Watson, J. M., & Moritz, J. B. (1999). The development of concepts of average. *Focus on Learning Problems in Mathematics*, *21*(4), 15-39.

Watson, J. M., & Moritz, J. B. (2000). The longitudinal development of understanding of average. *Mathematical Thinking and Learning*, *2*(1&2), 11-50.