Testing Young Children's Ideas of Mass Measurement

Jill Cheeseman
Monash University
Jill.Cheeseman@monash.edu.au

Andrea McDonough
Australian Catholic University
Andrea.McDonough@acu.edu.au

Abstract

This article reports an innovative use of photographs in a pencil-and-paper test which was developed to assess young children's understanding of mass measurement. Two hundred and ninety-five tests were administered by thirteen teachers of Years 1 and 2 children in 3 urban and rural schools. Many of these children of 6-8 years of age were able to "read" the photographs and diagrams in the test and recognise the images as representations of their classroom experiences. While the test had its limitations it also had some worthwhile features: it required explanation, deductive reasoning, and justification of thinking through the open response questions. We have demonstrated that it is possible to develop pencil-and-paper tests that use photographs and diagrams to closely connect written assessment with classroom experiences of young children. We believe that this study is important because it shows that such assessment tools can reveal a range of children's thinking and can be a useful addition to the various authentic assessment practices teachers adopt.

In a feature journal presenting research on learning, teaching, and using measurement, Smith, van den Heuvel-Panhuizen, and Teppo (2011) called for "curriculum developers [to] design more potent materials, teachers [to] teach the measurement content more effectively, and assessment professionals [to] develop more revealing assessments of learning" (p. 617). Following the teaching of a unit of work on the measurement of mass, we developed and administered a pencil-and-paper test to 6-8 year old children. This article reports on the trialing and evaluation of that assessment instrument.

Assessing young children

While pencil-and-paper tests have become popular because they are thought to be efficient, low-cost ways of assessing student thinking, there are also major criticisms of such tests:

> they could not tap students' ability to estimate the answer to arithmetic calculations, to construct geometric figures, to use calculators or rulers, or to produce complex, deductive arguments. (Smith & Stein, 2011, p. 155)

> [they] provide little evidence of a candidate's skills in solving complex problems in context, undertaking investigations, or carrying out particular practical mathematical tasks such as estimating distances in a local context, measuring using a variety of units, and manipulating space. (Izard & Miller, 1997, as cited in Ellerton & Clements, 1997, p.160)

For many years mathematics educators have been advocating more authentic methods of assessing mathematical learning (Ball & Bass, 2000; Clarke & Clarke, 2004; Leder, 1992; McKenney & Reeves, 2012). However, "externally written" pencil-and-paper tests, that is, those not written by the teachers themselves but prepared by people outside of the school, still comprise part of a range of assessment tools used in primary schools. One of the reasons pencil-and-paper tests continue to be used may be due to the move by educational sectors to increase "accountability" of teachers (Lowrie & Diezmann, 2009). The development of the test reported in this article was prompted by teachers' need to fulfil their school requirements for "topic" assessment which often took the form of pre- and post-evaluations. In fact, this requirement stimulated us, as researchers, to consider whether we could design a pencil-and-paper-test that was externally written and was authentic, open assessment that offered some insights into children's thinking.

Evaluating young children's mathematical thinking is usually not done with pencil-and-paper tests. Such tests involve abstract ideas interpreted through words, diagrams and symbols. It is hard for children to interpret the questions and to understand what they are required to do in response. The main reasons written tests are considered inappropriate assessment tools for young children concern the reading and writing difficulties they present for children of 6-8 years of age. These difficulties are not confined to young children (White, 2005) as over half the errors made by upper primary and early secondary students in Newman's (1977) seminal study were associated with what the author termed Reading, Comprehension or Transformation problems.

Keeping the reading and writing issues in mind, the language of our test was kept as simple as possible and constructed in short sentences. One way of trying to minimize reading difficulties was to specify that the test could be read aloud by the teacher. If necessary, the teacher could also ask a child whose writing was indecipherable to say what they had written as an answer to a test question. In the delivery of the test, 5 of the 13 project teachers elected to read the test and three supported the children's reading. A further two teachers annotated children's responses to make them legible. Teachers helped children to understand what they were being asked to do by the test questions and helped to scribe children's thinking.

Diagrams are another element of pencil-and-paper tests which are known to be difficult for students to interpret (Smith et al., 2011; van den Akker, Gravemeijer, McKenney, & Nieveen, 2006). As a result, the authors made every attempt to use photographs and diagrams which would be easily recognizable to the children and which would remind them of familiar classroom activities.

Another difficulty for children is that they often see tests as disconnected from their mathematical experiences. Older children who are more practiced in doing pencil-and-paper tests, for example, Year 3 Australian students who need to sit the National Assessment Program: Literacy and Numeracy (NAPLAN) often practice test questions in preparation. They rehearse various test question formats; they polish skills and are taught new facts in an effort to improve their test scores. However, for most students the tests are disconnected from their regular mathematical lessons and their experiences generally. The test reported here was designed to connect to children's experiences of mass measurement by using photographs and diagrams of known objects.

Of course as an assessment of knowledge and skills this pencil-and-paper test is limited in its scope as a tool to reveal mass measurement concepts. However, children's responses to open format test questions which required explanation and justification of their reasoning have provided us with some interesting data which give insights into children's thinking as discussed below.

In designing this test we were informed by Stenmark (1989) who stated that excellent mathematics assessment:

- matches the ideal curriculum in both what is taught and how it is experienced, with thoughtful questions that allow for thoughtful responses;
- communicates to students, teachers, and parents that most real problems cannot be solved quickly and that many have more than one answer;
- allows students to learn at their own pace\focuses on what students know and can do rather than what they don't know;
- won't label half the students as failures;
- doesn't use time as a factor, since speed is almost never relevant in mathematical effectiveness; and
- is integral to instruction and doesn't detract from the students' opportunities to continue to learn (p, 4).

It is with these attributes in mind that the mass measurement test was designed and that we later in this article reflect upon its overall value as an assessment tool.

Research context

The research project titled, *Investigating Early Concepts of Mass* (Cheeseman, McDonough, & Ferguson, 2012; McDonough, Cheeseman, & Ferguson, 2012a) was an attempt to meet the challenges related to the teaching and assessment of measurement put forward by Smith, van den Heuvel-Panhuizen, and Teppo (2011). The first phase was a teaching and learning research design project which offered teachers and children opportunities to take part in interesting lessons involving the measurement of mass (McDonough, Cheeseman, & Ferguson, 2012b). The project investigated productive ways to teach the measurement of mass with 6-8 year old children. A unit of lessons was developed and evaluated. These lessons progressed quickly from those focused on the use of informal units of mass to those transitioning children's thinking to formal units of measure. Having established the content of the lessons, we trialled them in Year 1 and Year 2 classrooms and evaluated the results. Teachers' views were sought and student learning was appraised using pre- and post-teaching one-to-one task-based interviews (Cheeseman et al., 2012).

In the second phase of the research, which we draw upon for this article, we worked with 3 urban and rural schools and their 13 Years 1 and 2 teachers and 295 students. These teachers taught the documented unit of lessons, supported each other and reflected on their experiences and the learning of the children. Time and the cost constraints meant that no interviews were undertaken to evaluate the student learning in this phase of the project. However, some evaluation of the learning of the children was considered important. As Clarke et al. (2002) noted:

> Most early years mathematics teachers would agree that their major aim in teaching mathematics is to build knowledge, skills and understandings, while engendering a positive attitude to the subject and enhancing students' own positive feelings about themselves as learners of mathematics (p. 44).

Indeed these were the major aims of the project and in order to assess the success of the lessons and the unit of work overall, it was important to have some evaluation of student knowledge, skills, understandings, and attitudes.

Teachers in the project observed children's learning and kept journals recording their observations and reflections. At our suggestion, several teachers used a pre-assessment task *Impress Me* where children were told "I want you to 'impress me' with all you know about mass and weighing." All of the children were asked to do the same task at the end of the unit of work. Analysis of these data can be found in McDonough and Cheeseman (in preparation). In addition, in response to a request from one of our project schools for an assessment tool, we provided some performance assessment tasks and we designed a pencil-and-paper test which is the subject of this article.

As reported in earlier papers regarding the teaching of the unit of work on which this test is based (e.g., McDonough et al., 2012a, 2012b), the learning experiences we developed for children were grounded in a philosophy of social constructivism, that is, in the belief that knowledge is individual and personal, socially constructed, and dependent on physical actions and interactions with others (e.g., Cobb, Wood, & Yackel, 1991; Ernest, 1994). We believe also that physical activities of measuring are underpinned by conceptual activity related to the development of foundational ideas of measure (e.g., Lehrer, Jaslow, & Curtis, 2003; Stephan & Clements, 2003; Wilson & Osborne, 1992). Such key ideas include *comparison*, that is, that "like properties can be compared to see which is greater" (Wilson & Osborne, 1992, p. 92) which for mass measurement can occur by holding objects, sometimes referred to as hefting, or by use of a scale such as a balance or digital scale. The concept of *unit* relates to key ideas including iteration and identical unit (Lehrer et al., 2003; Stephan & Clements, 2003) and is developed in relation to both informal units of mass such as plastic teddies or metal washes and formal or standard units including grams and kilograms. As had occurred for the related learning activities, the development of the test questions reflected our philosophical stance as well as our awareness of the importance of children coming to understand foundational ideas of measure.

Method

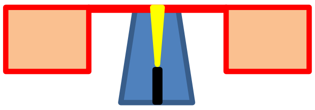
Two hundred and ninety-five pencil-and-paper tests were administered by 13 teachers of Years 1 and 2 children in 3 urban and rural schools. Teachers were authorised to administer the test in the way that they thought best suited the needs of their children. Five teachers read the test aloud to the children. A further five gave the paper to the children, read the first question aloud and left the children to go on at their own pace. If an individual child requested it, the teacher then read the question to that child. Three teachers left the children to read and answer the questions and one of these teachers wanted to see what the children could do without any support. The variation in delivery, not surprisingly, led to a variation in the completion of the test between class groups. The highest rates of missing data were from the two of the classes where no reading of the questions was offered. In the class where the teacher did not read the test aloud or offer reading support 7 children, of the 18 in the class, left the last half of the paper blank.

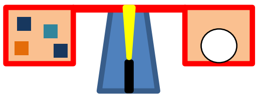
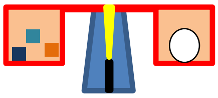
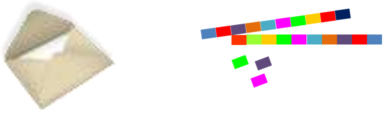
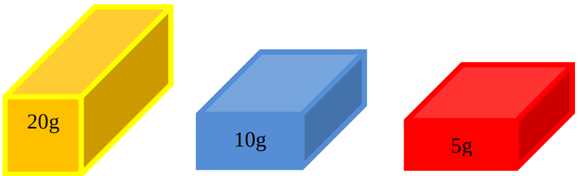
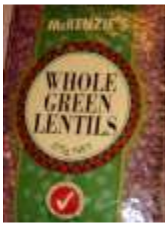
The test papers were read by the authors and responses were coded. Apart from correct responses, categories of code related to the thinking and reasoning of the children emerged from the data. An iterative process was used to capture the range of responses and the coding team checked with the first author for consistency of meaning and interpretation of children's responses. Twenty per cent of the sample was double coded and assigned "consensus codes" The overall inter-coder reliability was 79%. Data were entered into SPSS for analysis.

The test instrument

To provide a context for the results and findings that follow, the test questions and the mathematical thinking they were designed to address are shown in Table 1.

Table 1.
The Mass Measurement Test Questions Matched to their Intended Purpose.

Test question	Actions elicited/purpose
1. These hands are hefting. Draw a circle around the thing you think would feel heavier.  Why? ………………………………………………………	Judges the likely result of hefting two objects of which the children have some experience. Justification of decision.
2. Circle the box of teddies that would be heavier. Why? ....  	Understands that the more of a uniform unit you have the more they will weigh. Expresses the idea.
3. Some things are hidden in the buckets of this balance scale? What can you say about the weight of the things? …………………………………………………………… 	Notices the buckets are even and infers that the objects must be equal in mass.
4. Circle which is heavier in each scale. How do you know?  ……………………  ……………………  ……………………	Understands that the lower bucket holds the heavier mass. Notices that the same object on left side is heavier. Notices that similar size objects right side heavier. Notices larger object left side heavier. Three correct answers indicate consistent judgement of visual interpretation of balance scales.

5. Colour the ball that is heavier.   How do you know? ……………………………………………	Interprets the diagram. Understands the use of informal units to compare masses. Justifies thinking.
6. One Centicube weighs 1gram. This parcel was weighs the same as these Centicubes. How much does it weigh?  ……………………………	Can use the transition to formal unit ideas introduced in the lessons. Uses notation in formal units (grams).
7. This spoon is evenly balanced with four 10 g masses. How much does it weigh?  ………………………	Measures mass in formal composite units. Finds the total, records the number and the unit.
8. How much do these weights make together?  ………… grams	Adds formal masses based on abstract diagrammatic information.
9. Which of these packages weighs more? How can you tell? ……………………………………………………………   	Interprets the photos. Compares symbolic written masses.
10. The digital scales were used to weigh this lemon. How heavy is it?  ………………………..	Reads and interpret digital scales.
11. About how much do you think you weigh? ……………………………………………	Able to make a sensible estimate of, or recall, body weight.

Results

Data were analysed for frequency of correct response as summarised in Table 2.

Table 2.
Summary of Frequency of Correct Response to each Test Question.

Question number	1	2	3	4a	4b	4c	5	6	7	8	9	10	11
Facility (%)	84	99	86	96	97	96	52	53	54	73	61	63	21

The high facility with the first 4 test questions was notable. It indicates that children could interpret photographs and diagrams which closely related to their lived experience. The responses show that children could make reasoned judgements with regard to comparisons of masses and that they could interpret images of balance scales showing both equal and unequal masses. Young children attempted questions which involved several steps of thinking and deductive reasoning. Many could also understand questions that involved the use of formal units of mass measurement. Children's responses to each question were quite revealing of their thinking therefore a detailed analysis of each question follows.

Question 1. Children could make a reasoned interpretation of the picture of hefting (83%). Children used various strategies to do so: visual cues (37%), observation of the amount of material (24%), the nature of the materials (18%), and personal knowledge of the objects (14%). Only 7% were unable to explain their reasoning. Children demonstrated by their responses that they could infer ideas of mass by considering the look of objects, taking into account what they knew of the material of the objects.

Question 2. Almost all of children (97%) could interpret the photograph and reason about which box of teddies would weigh more. Of these, many (82%) gave a general response, similar to that in Figure 1 and a further 11% reasoned explicitly about 6 being more then 4 and so are heavier.

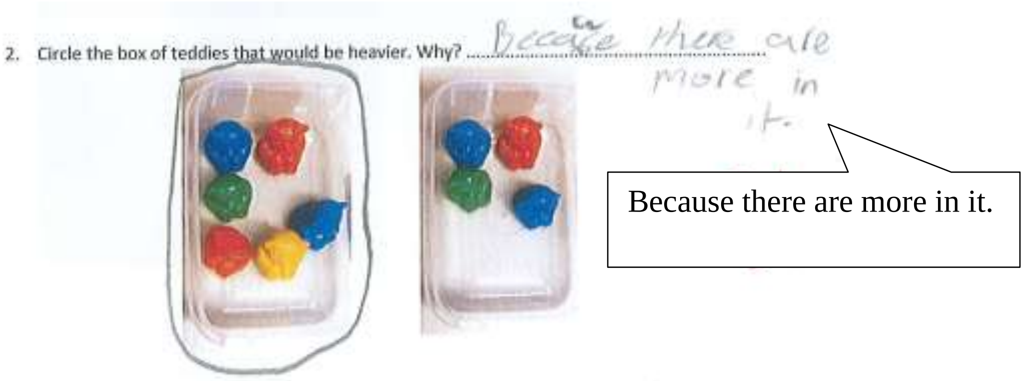


Figure 1. A sample of a general response.

Question 3 This question was thought by teachers to be quite difficult. The children had to interpret the diagram, to notice that the buckets on the balance scale were even and infer that whatever was hidden in each would be of equal mass. The majority of children (80%) understood the sides of the balance were “the same” (see Figure 2).

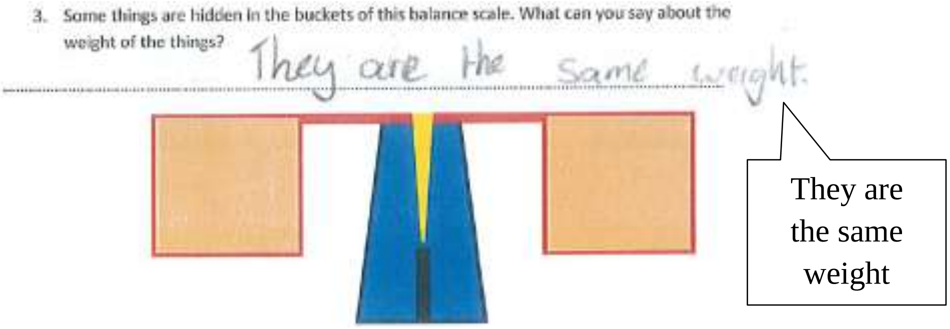


Figure 2. An example of children’s responses that interpreted the diagram as evenly balanced by equivalent masses.

Interestingly a further 17 (6%) children imagined exactly what might be in the buckets, for example, “A Papa bear on one side and another Papa bear on the other”. The responses sometimes mentioned identical object but others, such as Figure 3, illustrated an understanding that the objects in the buckets could be different and yet be the equivalent masses.

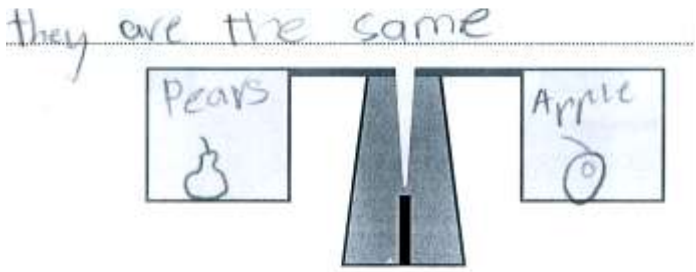
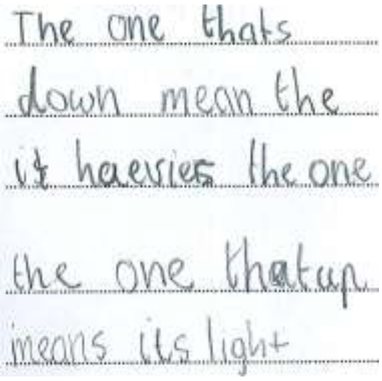


Figure 3. An example of a response imagining objects which would balance as they were equivalent masses.

In debriefing meetings after the tests were administered teachers specifically mentioned this question as being “tricky” for the children. One teacher reported that although he had said nothing, he placed a balance scale in plain sight of the children and used his hands to cover the buckets. His class results were at the top of the range of facility with this question. This teacher’s comments call into question the results for this question and they raise the general issue of teachers not really expecting children to struggle with test items which involve interpretation, reasoning and deduction. Taken at face value though, 86% of children could correctly read the diagram and interpret the scale as balanced. In addition some of the responses were very sophisticated as shown in Figure 4.

Figure 4. A response that shows a child imagining equal masses in standard units.

Question 4. Each of the three parts of this question involved children reading a photograph of unbalanced scales, deciding which object was heavier, and explaining the reason for their choice. Results were consistently high see Table 2 with many children expressing ideas similar to those in Figure 5.

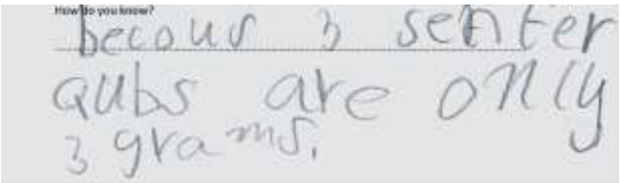


The one that's down means it is the heaviest. The one that's up means it's light.

Figure 5. A common explanation of reading the balance scales.

The justification children gave for their decisions fell into 3 main categories. As shown above many looked at the "down side" of the scales, some children used visual cues and size judgments e.g., "it looks bigger". Others used reasoning based on the fabric of the object e.g., "the calculator is made of plastic" or their knowledge of the object e.g., "Because a mug is heavy".

Question 5. Just over half (52%) of children could correctly decide which ball was heavier in the diagram in Question 5. About one third of these children used a strategy of comparing the number of blocks in the balance. One particularly insightful response is shown in Figure 3 in which the child interpreted the cubes in the diagram as materials she had handled and used (Centicubes) to then reason that one of the balls weighed 4 grams and the other ball weighed 3 grams so she correctly coloured the heavier one.



Because 3 Centicubes are only 3 grams

Figure 6. Reasoning about the comparison of masses of the balls.

The incorrect reasoning of children was hard to understand in 20% of responses and a further 19% compared the blocks incorrectly.

Question 6.

This question was developed to reflect the experience of the children in connecting Centicubes in 10 g composite masses and using them to accurately weigh post parcels. To be correct the response needed to have the number and the unit (in grams). More than half (52%) of the children were completely correct and an additional 18% gave the correct number only.

This response may be an indicator of knowledge of counting rather than any connection to mass measurement.

Question 7.

The pictured plastic commercial weights were available to children in class. The photograph of measuring mass in formal composite commercial units was interpreted correctly by more than half (54%) of children who recorded the number and the unit. As in the previous question, a further 13% of the responses correctly calculated the number but recorded no unit. This raised queries about the extent to which children were focused on the creation of a total mass and assumed the unit as grams. Only during one of the five lessons was the formal notation of mass presented to the children other than incidentally.

Question 8. This question required the addition of different composite units of mass and 73% of children could correctly calculate the sum.

Question 9.

Of the 61% of children who correctly identified the tin of tomatoes as the heaviest, nearly half said they read the label e.g., "it tells you on the packet." One fifth (20%) of total number responses gave a reason based on the characteristics of the contents. This thinking was more prevalent in the incorrect responses where more than half (61%) reflected this reasoning. Children considered the materials from which objects were made e.g., the child selected the packet of lentils "Because it has more sturf [*sic*] in it" and their reasoning for a response that circled the beans being, "The nuts because there's lots in it." These responses seem to be based on the number of things inside the container. However, in science education research literature mass and weight are discussed in the context of concepts of matter – in children's understanding of substances (Papageorgiou & Johnson, 2005). Perhaps in this context the thinking has more meaning.

Question 10.

Result show that children can accurately read digital scales in a photograph (91%). Of these children 63% were completely correct and again 28% wrote a number with no unit. We hypothesise that because the children had little formal exposure to mass notation and because the sequence of lessons only used grams as formal units the children assumed that we were measuring in grams.

Question 11.

This was the only question on the test with which the children had no classroom experience. Teachers were amazed that 21% of their children either knew a good estimate or knew their exact body weight in kilograms. An indication that perhaps the lessons had rehearsed children to think in grams was given by about ¼ of the responses which showed a number in a suitable range accompanied by the unit in grams e.g., "I weigh 43 g."

Discussion of results

Taking an overview of the findings has led us to consider 4 main themes. These are the:

- successful use of images, both photographs and diagrams;
- connection of the test questions with the classroom experiences;
- elicitation of reasoning; and
- revelation of children's emerging ideas of mass measurement.

Photographs and diagrams.

One of the innovative elements of the test was the use of images throughout. Much has been written about the difficulties children have interpreting mathematical diagrams (Lowrie, Diezmann, Logan, 2012), however, the results reported here indicate that when diagrams and photographs are closely connected to children's mathematical experiences interpretation difficulties are minimised.

Connection to experience.

The test questions mapped onto a series of 5 mass lessons the children had experienced. It used images that were intended to be familiar to children and to remind them of the equipment they had handled and explored. It was clear from the responses of the children and the comments of their teachers that these connections were plain. When asked to evaluate the test one teacher remarked, "I loved the way the test used pictures of the things we did. The children were saying things like, Oh yes, I remember, we weighed things with those teddies!" Children also connected the test images to their experience in and out of the classroom, for example, one child wrote, "I think the cup is the heaviest because when I have hot chocolate it is really heavy."

Mathematical reasoning.

The expectation of children to explain and justify their thinking was part of the test design. The questions asked children: Why? How do you know? What can you say about …? How can you tell? Many children recorded their mathematical reasoning and gave us insights into their thinking about mass measurement and their logical and deductive reasoning in general.

We will use one child's response to illustrate how these features were seen. Figure 8 illustrates a child's interpretation of a diagram, his connection to classroom experience, and his mathematical reasoning with dynamic thinking. We read this response as the child anticipating what will happen as a cartoon sequence strip, along these lines: if 4 blocks balance the ball, then 3 blocks will go up and the ball will go down. While the response was coded incorrect, it gives a real insight into the child's interpretation of the diagram, his connection to his experiences with balance scales and his visualisation of the action.

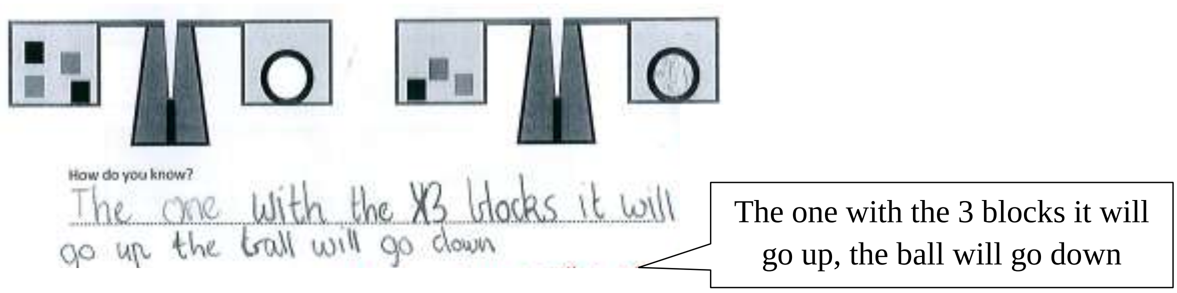


Figure 8. A response that shows a child anticipating a future action.

Children's emerging interpretations of balance scales

Children's emergent concepts of mass measurement have been detailed and discussed elsewhere (McDonough et al., 2012a). Their identification was the result of classroom observations by the authors. We were intrigued to find some similar ideas in the responses on a pencil-and-paper test. For example, children offer explanations related to the material of the object rather than reasoning based on the position of the balance scale when deciding which object is heavier. While we concede that this may be a common sense thing to do to consider the characteristics of the object that is being compared, we hypothesise that some 6 and 7 year-olds do not yet "trust the scale". Perhaps for these children the bucket of the balance being down is not as convincing as knowing that "a cup is made of glass" so it is heavier. Researchers in early counting (Cowan, 1987; Treacy & Willis, 2003) have identified a phase when the child does not "trust the count" and understand that no matter what way they count a collection they get the same result. Once children trust in the counting process they use it to solve relational problems. Perhaps in a similar vein of thinking the children whose thinking we are reporting here do not yet trust the balance scale. The way a balance scale works is perhaps still being conceived and until it is, "trusted" relational judgments are based on their knowledge of the object.

Interpreting a balance scale is perhaps more complex than we realise. Asked to circle the heavier object in Figure 8 and to justify the answer one child said, "I can't tell because it is not in properly".

Figure 8. Photograph showing two unequal buckets of a balance scale.

In a different context we have observed children judge mass by comparing the upper edge of objects in the buckets. We have also noticed children fixing their attention on pointers and beams of the balances. In general we are conscious that further research is needed to understand exactly what children are attending to when they look at balance scales.

A final reflection on the assessment protocol

In evaluating the test as a whole we return to the attributes specified by Stenmark (1989) and take each in turn as a stimulus for reflection. This test fulfils Stenmark's aim that assessment "matches the ideal curriculum in both what is taught and how it is experienced, with thoughtful questions that allow for thoughtful responses". Many children of 6-8 years could "read" the photographs and diagrams in the test. They recognised the images as representations of their classroom experiences. The great majority of young children could respond to pencil-and-paper test questions dealing with comparison of masses. These comparisons could be reasoned through a combination of visual information and knowledge of objects and the materials from which they are made.

The test elicited concepts of mass measurement in more than half of the children. Questions involving informal and formal units were mastered by more than half of the children. This was shown by responses with a correct number and a correct unit. If we had been less exacting about the notation and accepted a correct number and an assumed unit of mass, approximately ¾ of the children would have been found to answer formal measurement questions correctly. It is notable that ideas of formal units of mass measurement go beyond the intended Australian Curriculum: Mathematics at Years 1 and 2 (Australian Curriculum Assessment and Reporting Authority, 2012). This raises Stenmark's aim of a match to the "ideal curriculum" and we would argue that the responses of these students are indeed ideal.

While the test had its limitations in that it did not require children to measure mass in context, it did have abstract representations of contexts which were familiar to children. It required explanation, deductive reasoning and justification of thinking through open response questions.

This test fulfils aims expressed by Stenmark (1989, p. 4) in that the test was an attempt to focus on what children "know and can do rather than what they don't know", the test did not "label half the students as failures" and it was designed to be "integral to instruction". The high proportion of correct responses and the students' comments indicating that they recognized the pictures as those related to their classroom experiences, suggest that these aims have been met. In addition the test communicated to teachers and students that questions could be posed as "real problems" similar to those they had tackled in their classroom activities.

The test did not "use time as a factor" as students were given ample time to complete the questions and some teachers provided reading support to assist their students to understand what was being asked of them in the test.

We have demonstrated that it is possible to develop pencil-and-paper tests that use photographs and diagrams to closely connect written assessment to classroom experiences of young children. Such assessment tools can reveal a range of children's thinking and can be a useful addition to the various authentic assessment practices teachers adopt.

References

Australian Curriculum Assessment and Reporting Authority. (2012). Australian Curriculum: Mathematics Retrieved 23rd January 2012, from http://www.australiancurriculum.edu.au/Mathematics/Curriculum/F-10

Ball, D., & Bass, H. (2000). Interweaving content and pedagogy in teaching and learning to teach: knowing and using mathematics. In J. Boaler (Ed.), *Multiple perspectives on learning and teaching* (pp. 83–104). Westport: Ablex.

Cheeseman, J., McDonough, A., & Ferguson, S. (2012). The effects of creating rich learning environments for children to measure mass. In D. Jaguthsing & C. L. Pien (Eds.), *Mathematics education: Expanding horizons*. Singapore: Mathematics Education Research Group of Australasia.

Clarke, D. M., Cheeseman, J., Gervasoni, A., Gronn, D., Horne, M., McDonough, A., . . . Rowley, G. (2002). *Early Numeracy Research Project: Final report, February 2002*. Fitzroy, Victoria: Australian Catholic University, Mathematics Teaching and Learning Centre.

Clarke, D. M., & Clarke, B. A. (2004). Mathematics teaching in grades k-2: painting a picture of challenging, supportive, and effective classrooms. In R. N. Rubenstein & G. W. Bright (Eds.), *Perspectives on the teaching of mathematics (66th yearbook of the national council of teachers of mathematics)* (pp. 67-81). Reston, VA: NCTM.

Cobb, P., Wood, T., & Yackel, E. (1991). A constructivist approach to second grade mathematics. In E. von Glasersfeld (Ed.), *Radical constructivism in mathematics education,* (pp. 157–176). Dortrecht, The Netherlands: Kluwer.

Cowan, R. (1987). When do children trust counting as a basis for relative number judgments? *Journal of Experiemental Child Psychology, 43*(3), 328-345.

Ellerton, N., & Clements, M. (1997). Pencil and paper tests under the microscope. In F. Biddulph & K. Carr (Eds.), *People in mathematics education* (pp. 155-162). Waikato, NZ: Mathematics Education Research Group of Australasia.

Ernest, P. (1994). Varieties of constructivism: Their metaphors, epistemologies and pedagogical implications. *Hiroshima Journal of Mathematics Education, 2*, 1-14.

Leder, G. E. (1992). *Assessment and learning of mathematics.* Hawthorn, Victoria: Australian Council for Educational Research.

Lehrer, R., Jaslow, L., & Curtis, C. (2003). Developing an understanding of measurement in the elementary grades. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement* (pp. 100-121). Reston, VA: National Council of Teachers of Mathematics.

Lowrie, T., & Diezmann, C. M. (2009). National numeracy tests: A graphic tells a thousand words. *Australian Journal of Education, 52*(2), 141-157.

Lowrie, T., Diezmann, C. M., & Logan, T. (2012). A framework for mathematics graphical tasks: the influence of the graphic element on student sense making. *Mathematics Education Research Journal, 24*(2), 169-187.

McDonough, A., Cheeseman, J., & Ferguson, S. (2012a). Insights into children's understandings of mass measurement. In T. Y. Tso (Ed.), *Proceedings of the 36th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 201-208). Taipei, Taiwan: PME.

McDonough, A., Cheeseman, J., & Ferguson, S. (2012b). *Striving to maximize children's learning of measurement.* Paper presented at the 12th International Congress on Mathematical Education, 8 July – 15 July, 2012, Seoul, Korea.

McKenney, S., & Reeves, T. (2012). *Conducting educational design research: What, why & how*. London: Routeldge.

Newman, M. A. (1977). An analysis of sixth-grade pupils' errors on written mathematical tasks. *Victorian Institute for Educational Research Bulletin, 39*, 31-43.

Papageorgiou, G., & Johnson, P. (2005). Do Particle Ideas Help or Hinder Pupils' Understanding of Phenomena? *International Journal of Science Education, 27*(11), 1299-1317.

Smith, J., van den Heuvel-Panhuizen, M., & Teppo, A. (2011). Learning, teaching, and using measurement: introduction to the issue. *ZDM, 43*(5), 617-620.

Smith, M., & Stein, M. K. (2011). *5 practices for orchestrating productive mathematics discussions*. Reston, VA: National Council of Teachers of Mathematics.

Stenmark, J. K. (1989). *Assessment alternatives in mathematics*. Berkley, California: Lawrence Hall of Science.

Stephan, M. & Clements, D. H. (2003). Linear and area measurement in prekindergarten to grade 2. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement* (2003 Yearbook, pp. 3-16). Reston, VA: National Council of Teachers of Mathematics.

Treacy, K., & Willis, S. (2003). A model of early number development. In L. Bragg, C. Campbell, G. Herbert & J. Mousley (Eds.), *Mathematics education research: Innovation, networking, opportunity. Proceedings of the 26th Annual Conference of the mathematics Education Research Group of Australasia* (pp. 674-681). Sydney: MERGA.

van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). (2006). *Educational design research*. New York: Routledge.

White, A. (2005). Active mathematics In classrooms: Finding out why children make mistakes-and then doing something to help them. *Square One*, *15*(4), 15-19.

Wilson, P. A., & Osborne, A. (1992). Foundational ideas in teaching about measure. In T. R. Post (Ed.), *Teaching mathematics in grades K-8: Research-based methods* (pp. 89-121). Needham Heights, Massachusetts: Allyn & Bacon.