Using Photographs and Diagrams to Test Young Children's Mass Thinking

Jill Cheeseman
Monash University
<Jill.Cheeseman@education.monash.edu.au>

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

This paper reports the results of a pencil-and-paper test developed to assess young children's understanding of mass measurement. The innovative element of the test was its use of photographs. We found many children of the 295 6-8 year-old children tested could "read" the photographs and diagrams and recognise the images as representations of their classroom experiences. While the test had its limitations, it also 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 to classroom experiences of young children. Such assessment tools can reveal a range of children's thinking and can be a useful addition to various authentic assessment practices.

There is continued recognition today that assessment is central to learning (e.g., Wiliam, 2010) and various assessment instruments ranging from national tests to daily formative assessments are used in classrooms in Australia (Santiago, Donaldson, Herman, & Shewbridge, 2011). However, in a feature journal presenting research on learning, teaching, and using measurement Smith, van den Heuvel-Panhuizen, and Teppo (2011) pointed to the need for further development of appropriate assessment tools in the area of measurement calling on, "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).

The research project entitled, Investigating Early Concepts of Mass (Cheeseman, McDonough, & Ferguson, 2012; McDonough, Cheeseman, & Ferguson, 2012a) was an attempt to meet these challenges. It 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. 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 we worked with three 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 children's learning. Time and the cost constraints meant that no interviews were undertaken to evaluate the student learning in this phase of the project. However, evaluation of the learning of the children was considered important in order to assess the success of the lessons and the unit of work overall. Teachers in the project observed children's learning and kept journals recording their field notes and reflections. They used a pre- and post- open ended assessment task and a pencil-and-paper test. It is analysis of responses to this test that is the subject of this paper.

Many decades ago Stenmark (1989) said we need mathematics assessment that:

> 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

In V. Steinle, L. Ball & C. Bardini (Eds.), Mathematics education: Yesterday, today and tomorrow (Proceedings of the 36th annual conference of the Mathematics Education Research Group of Australasia). Melbourne, VIC: MERGA.

> 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).

Long after they were first published, these attributes still characterise excellent mathematics assessment and are demanding criteria by which to judge assessment protocols. With these attributes in mind, a mass measurement test was designed and its results will be examined.

Pencil-and-paper tests

While pencil-and-paper tests have become popular because they are thought to be efficient and low-cost, 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 (Clarke & Clarke, 2004; Leder, 1992; McKenney & Reeves, 2012). However, externally written pencil-and-paper tests 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 here 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 the requirement stimulated us, as researchers, to consider whether we could design a pencil-and-paper-test that was "externally written" i.e. not by the teachers themselves, and was authentic, open assessment which offered some insights into children's thinking.

Assessing young children

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 (Newman, 1977; White, 2005).

Keeping the reading and writing issues in mind, the language of the test was kept as simple as possible and constructed in short sentences. One way of trying to minimize reading difficulties was to specify that it could 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 questions on the test and helped to interpret their 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 A ustralian 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 many 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 response formats that required explanation and justification of their reasoning on the test have provided us with some interesting data which give insights into children's thinking.

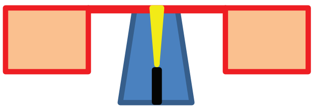
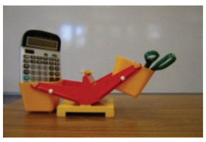
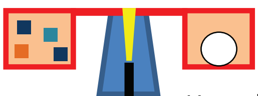
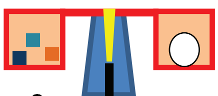
Method

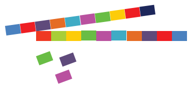
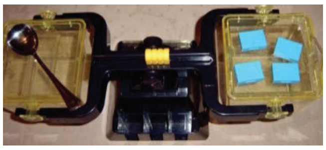
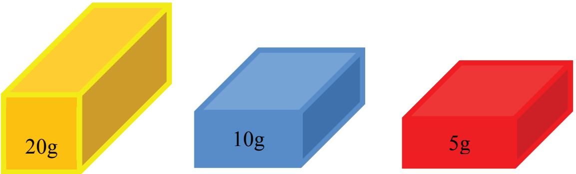
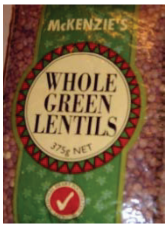
Two hundred and ninety-five pencil-and-paper tests were administered by 13 teachers of Years 1 and 2 children in three urban and rural schools. Teachers were authorised to administer the test in the way that they thought best suited the needs of their children. Five teachers read the test aloud to the children. A further five gave the paper to the children, read the first question aloud and left the children to go on a t 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. To provide a context for the results and findings which follow, the test questions, the percentage of children who gave correct responses for each question, and the mathematical thinking each was designed to address are shown in Table 1.

Table 1

The Mass Measurement Test Questions their Facility and 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? ………………………………………	84	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? …………………………………………………… 	99	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? …………………………………………………………………………………… 	86	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?  ……………………  ……………………  ……………………	96 97 96	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? ……………………………	52	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?	53	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?  ………………………	54	Measures mass in formal composite units. Finds the total, records the number and the unit.
8. How much do these weights make together?  ………… grams	73	Adds formal masses based on abstract diagrammatic information.
9. Which of these packages weighs more? How can you tell? ………………………………………………………   	61	Interprets the photos. Compares symbolic written masses.
10. The digital scales were used to weigh this lemon. How heavy is it?  ………………………..	63	Reads and interpret digital scales.
11. About how much do you think you weigh? ……………………………………………	21	Able to make an estimate of, or recall, body weight.

Results and discussion

Children's responses to each question have been detailed elsewhere (Cheeseman & McDonough, 2013). Taking an overview of the findings has led us to consider four main themes which are the focus of the discussion in this paper. 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 & Deitzman, 2012). However, these results indicate that when diagrams and photographs are closely connected to children's mathematical experiences interpretation difficulties are minimised. The high frequency of correct response (shown in Table 1) with the first four test questions was notable. It indicates that children could interpret photographs and diagrams closely related to their lived experience. A detailed analysis of responses showed 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.

Connection to experience.

The test questions mapped onto a series of five mass lessons the children had experienced. The test 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 which gave insights into their thinking about mass measurement and their logical and deductive reasoning in general. One child's response illustrates these features. A child's interpretation of a diagram (Question 5) shows his connection to classroom experience, and his mathematical reasoning with dynamic thinking. He wrote, "The one with the 3 blocks it will go up, the ball will go down." 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.

In response to the same question: How do you know? another child's response was particularly insightful. She wrote, "Becos 3 centercubs (sic) are only 3 grams". She had handled and used materials (cubic centimetre blocks with the mass of a gram) then used her experience and knowledge to reason correctly that one of the balls weighed four grams and the other weighed three grams.

Children's emerging interpretations of balance scales

Children's emergent concepts of mass measurement have been detailed and discussed elsewhere (McDonough, Cheeseman & Ferguson, 2012a). Their identification was the result of classroom observations. 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 reason 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, we hypothesise that some 6 and 7 year-olds do not yet "trust the scale". Perhaps for these children the position of the bucket of the balance is not as convincing as knowing that "a cup is made of glass". Researchers in early counting (Cowan, 1987; Treacy & Willis, 2003) identified a phase when the child does not "trust the count" and understand that no matter which 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, some children whose thinking we are reporting here do not yet trust the balance scale. The way a balance scale works is possibly still being conceived and until it is "trusted" relational judgments are based on their knowledge of the object instead.

Interpreting a balance scale is perhaps more complex than we realise. Asked to circle the heavier object in Question 4a and to justify the answer one child said, "I can't tell because it is not in properly". In classrooms we have observed children judge mass by comparing the upper edge of objects in the buckets. We have also noticed children attending to pointers and beams of balances. In general we are conscious that further research is needed to understand exactly what children are noticing when they look at balance scales.

A final observation

Many children of 6-8 years could "read" the photographs and diagrams in the test and recognise the images as representations of their classroom experiences. The great majority of young children could respond to 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. Judgements could also be made by interpreting balance scales. 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. Had we 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.

While the test had its limitations - it did not require children to measure mass, it did have abstract representations of contexts which were familiar to children. It used open response formats to require explanation, deductive reasoning and justification of thinking of young children. We claim this test fulfils aims expressed by Stenmark (1989) because it "matches the ideal curriculum in both what is taught and how it is experienced, with thoughtful questions that allow for thoughtful responses". The test was an attempt to focus on what children "know and can do rather than what they don't know" and it was designed to be "integral to instruction" (p. 4). 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 teachers' authentic assessment practices.

References

Cheeseman, J., & McDonough, A. (2013). Testing young children's ideas of mass measurement. *International Journal of Mathematics Teaching and Learning.*

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., & 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.

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.

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

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

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.

McDonough, A., & Clarke, D. M. (2003). Describing the practice of effective teachers of mathematics in the early years. In N. A. Pateman, B. J. Dougherty & J. T. Zilliox (Eds.), *Proceedings of the 2003 joint meeting of the International Group for the Psychology of Mathematics Education and the Psychology of Mathematics Education Group North America* (Vol. 3, pp. 261-268). Hawaii: University of Hawaii.

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.

Santiago, P., Donaldson, G., Herman, J., & Shewbridge, C. (2011). Student assessment. In *OECD Reviews of evaluation and assessment in education: Australia 2011.* OECD Publishing. http://dx.doi.org/10.1787/9789264116672-6-en

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.

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.

Wiliam, D. (2010). The role of formative assessment in effective learning environments. In H. Dumont, D. Istance & F. Benavides (Eds.), *The nature of learning: Using research to inspire practice.* OECD Publishing. http://dx.doi.org/10.1787/9789264086487-8-en