DOCUMENT RESUME

ED 368 781 TM 021 294

AUTHOR Doig, Brian
TITLE Curriculum-Embedded Assessment in Mathematics.
PUB DATE Apr 94
NOTE 17p.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Educational Assessment; Elementary Secondary Education; Foreign Countries; Mathematics Achievement; *Mathematics Tests; Scoring; Standardized Tests; *Standards; *Test Construction; *Testing Problems
IDENTIFIERS Australia; *Australian Council for Educational Research; Calibration; *Curriculum Embedded Tests; Performance Based Evaluation

ABSTRACT

The curriculum-embedded procedures used to construct, validate, and refine assessment tasks for mathematics are described and discussed. Curriculum-embedded assessment places assessment tasks in the day-to-day context of the classroom. The test of a curriculum-embedded task is whether it could be regarded as curriculum material per se. In Australia, as in other countries, national standards for student achievement have been constructed and published. Any set of assessment tasks linked firmly to these standards should provide standardized reporting. Standardization of curriculum-embedded alternative assessments is possible if attention is paid to problems of test administration and scoring during test development. A bank of teacher-selected and teacher-administered tasks standardized to enable system-wide reporting and scored on a partial credit basis would enable formative assessments to be made. To conform to the ground rules of curriculum embedded assessment the Developmental Assessment Resource for Teachers (DART) project of the Australian Council for Educational Research has developed activities for lower and upper grades. Calibration of these activities will be conducted with a national sample of Australian students. Eleven figures provide examples of the DART activities. (Contains 2 references.) (SLD)

Curriculum-embedded assessment in mathematics

Brian Doig

The Australian Council for Educational Research

The intention of this paper is to outline the procedures used to construct, validate and refine assessment tasks for upper grade mathematics which are classified as curriculum-embedded.

Curriculum-embedded assessment places assessment tasks within the day-to-day context of the classroom, and whilst such tasks are essentially for assessment, they necessarily have strong curriculum roots. The test of a curriculum-embedded task is whether it could be regarded as curriculum material *per se*.

While school administrators need reliable summative data on student performance, teachers need formative information. In many cases these two needs are at odds with one another. However tasks that provide formative information for teachers can also provide 'standards' information for administrators.

Typical characteristics of existing assessment materials are that they are pencil-and-paper, usually with a single correct response within a multiple choice format. Assessment are conducted in silence with individuals working alone for a specified time. In contrast with this day-to-day classroom activities usually employ manipulative materials, verbal responses, discussion, and group work. The time allotted to these activities varies, and there may be more than one correct answer; indeed there may be a focus on the methods used to solve the problem rather than an answer to the problem itself. Curriculum-embedded assessment must attempt to reflect these latter, classroom characteristics, and be as un-intrusive as possible. It is suggested then that curriculum-embedded assessment must incorporate:

more than pencil-and-paper tasks;
a range of answers to be scored;
no all-or-nothing (right/wrong) scoring;
matching of the task to the child;
providing individual students with their own set of tasks;
allowing different tasks to assess the same ability;
reporting in a manner similar to traditional assessment forms.

A major issue in any assessment practice is management. A definite advantage of traditional assessment practice is that everyone doing the same set of items reduces administration and scoring time; queries about word meanings are easily handled for everyone at once; parents and administrators are satisfied that results are reliable due to the common items and standardised scoring. Any alternative assessment must attempt to provide as few new management problems as possible. Two main problems of management of the type of assessment being suggested both stem from providing students with different sets of tasks to complete. Managing twenty students who are involved in several different tasks would be a nightmare unless the tasks do not have some standard form. The classroom activities described above do use a recipe format and it seems sensible to follow this pattern. Once one assessment task has been completed, sufficient knowledge of the format should be gained to enable students to be independent of the teacher., so minimising teacher's management problems.

The second aspect of management that is problematic with alternative assessments of the type envisaged is their face validity with respect to acceptance as being fair to students. That is, if different tasks are set for different students, how can comparisons be made or grades given? Essentially, being fair is both a statistical and reporting problem.

As with any assessment item, calibration is essential. Every task must have a known difficulty estimated independently of the students who attempted the item. Item difficulties need to be robust enough to give us confidence that our own students need not reflect identically the group with which these items were calibrated in the first place. In any given domain of interest, an item assessing that domain represents an instance of that domain, and there are an infinity of other instances. A student's ability, estimated from their successes and failures on items, must be independent of the particular tasks they undertook. This is of course true for traditional assessment instruments as well. Independence tells us that it is feasible for students to attempt different tasks and yet be assessed on the same domain.

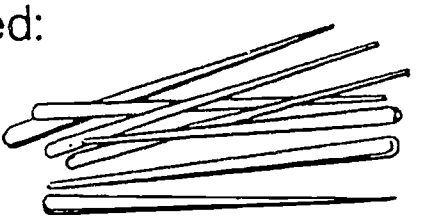
Reporting the performances of students on assessment tasks is straightforward when all students do the same items. In the case of curriculum-embedded assessment, the curriculum upon which the assessment tasks are based provides the beginning of a frame of reference for reporting. In Australia as well as other countries, national standards for student achievement have been constructed and published. The Australian 'Profiles in mathematics' are described as a framework for assessment and reporting (AEC, 1991). This being the case, any set of assessment tasks firmly linked to these profiles should provide standardised reporting. Standardisation of administration and scoring is possible if attention is paid to these aspects during the development stage. In order to provide formative information for teachers, simple right/wrong scoring of children's performance is not useful. Scoring should give information about children who fail to complete fully and successsfully any task. Scores need to be assigned to partial answers and such partial credit scores used for reporting on progress; in this way teachers gain formative information whilst summative assessment is being conducted.

In essence the result would be a bank of teacher selected and administered tasks, standardised to enable system-wide reporting, and scored on a partial credit basis for formative assessments to be made. Teachers can further gain because tasks may be matched to individual student needs whilst administrators gain because the information gathered is standardized and the reporting of results is within a fixed framework. The primary aim then is to create valid, reliable, standardised assessment tasks in a format that embedded the tasks in day-to-day classroom practice, which allow teachers to select any set of such tasks for administration to any single child or group of children. The necessity for whole class testing has to be removed and assessment tasks must look and feel like normal classroom activities. The ground rules adopted for creating a usable collection of curriculum-embedded tasks were:

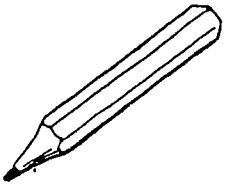
normal classroom look and feel;
user-friendly style to encourage children;
teacher choice of appropriate tasks;
any set of tasks could be used for a standardised assessment;
formative information to be provided from task performance;
standardisation of results from specified scoring criteria;
tied to a standard reporting framework;

You will need:

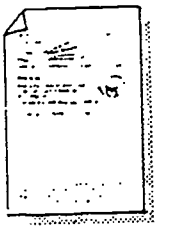
plasticine



toothpicks

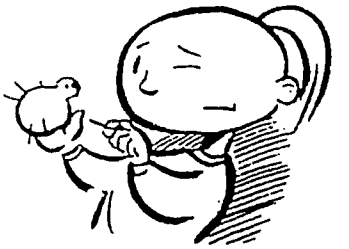


a pencil



a sheet

What to do



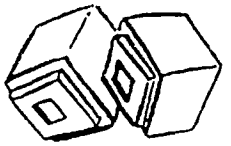
Make some creatures which have 6 legs using balls of plasticine and toothpicks.

Complete this table about your creatures.

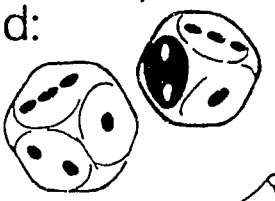
Creatures	Bodies	Legs
1	1	6
2		
3		
4		
5		

Estimate how many bodies and legs you would need to make 20 creatures. How do you know?

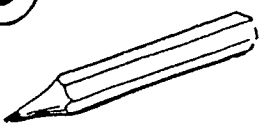
You will need:



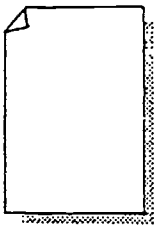
Unifix cubes



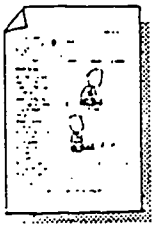
dice



a pencil



paper



a sheet

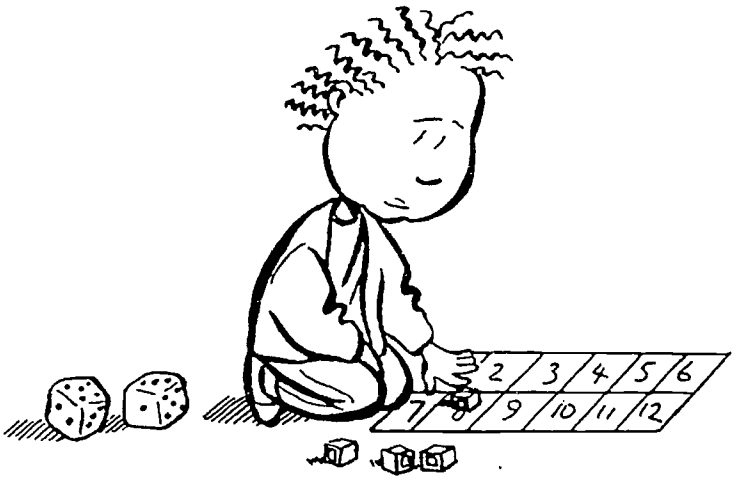
What to do

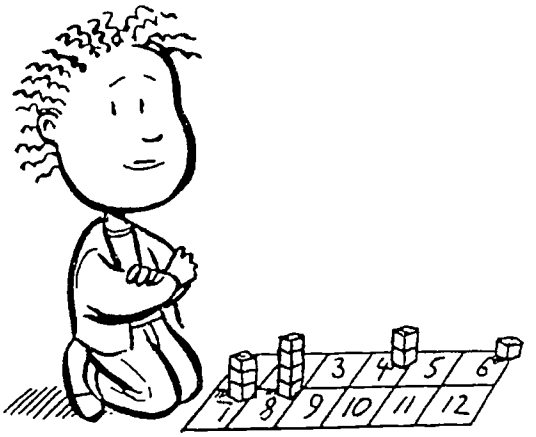
Roll the dice and place a Unifix cube on that total on the grid.

Keep rolling the dice, say 50 times, to build towers on the grid.

Make a poster showing:

- how many rolls you had altogether
- how many towers you made
- which tower was the tallest
- which tower was the shortest.

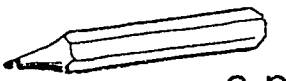




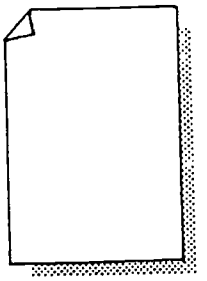
	2	3	4	5	6
7	8	9	10	11	12

Try again and see if you get similar results.

You will need:

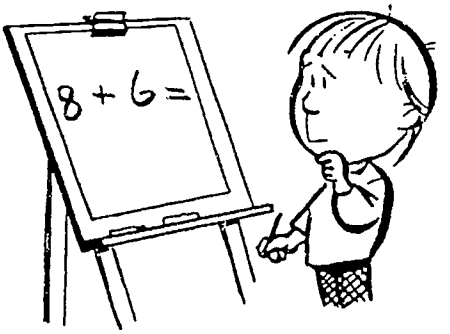


a pencil



paper

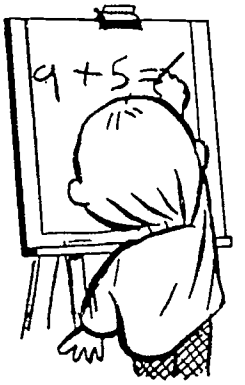
What to do



Choose any two even numbers. Add them. Is the total odd or even?

Add two odd numbers. Is the total odd or even?

Add an odd and an even number. Is the total odd or even?



Try these adding patterns a few times with different numbers.

Can you find a pattern?

Make a poster to show what you've discovered.

An attempt to produce assessment tasks that conformed to the ground rules above led to the commencement of the Developmental Assessment Resource for Teachers (DART) project at the Australian Council for Educational Research. Since DART was intended to reflect the child's learning environment DART activities would be indistinguishable from the class's day-to-day curriculum activities, and DART activities were to be learning oriented. The teacher's freedom to select child appropriate DART activities means that there would need to be several DART activities for each outcome of learning defined by the national Profiles, and any selected set of DART activities would constitute a reliable assessment.

Where to begin? The first stage was to gain an overview of current classroom practice in terms of the type of activities presented to children. An examination of typical classroom activities (or worksheets) shows an enormous range, from the common drill and practice worksheet to the more adventurous activity sheets such as those below (activities 3, 7, and 11: Doig, 1989).

Features that make these activity sheets easy to use in the classroom are the standard layout, the recipe form of the instructions, and the use of graphics to complement the text. A child who has used one of these activities is usually able to manage any other independently of the teacher. This makes such activities manageable in the classroom, given that not all children would be using the same activity at the same time. Text is kept to a minimum, although there must be sufficient to make quite clear what is to be done, otherwise teachers will be either driven mad by requests for clarification or will resort to whole class usage.

The next step was to examine the national curriculum framework (commonly known as the 'Maths statement'); these divide the mathematics curriculum into six strands. These are Algebra, Chance and Data, Measurement, Number, Space and Working Mathematically. The first five represent the mathematical content areas, while the last is focused on mathematical processes. Each of the six strands spans the content of school mathematics from the first year of school until the end of compulsory schooling. These years are divided into eight levels for assessing progress. (That is, into 18 month portions). For example, level four indicates the expected achievements of a child at the end of their primary schooling. Despite the apparent curriculum emphasis of the profiles, they are in fact a framework for assessment and reporting children's mathematical achievements. Statements of achievement at each level summarize the mathematical outcomes that can be expected of children at that level. These outcome statements therefore can form both the focus of assessment and the means of reporting a child's achievements. A school may define its own curriculum by aligning its learning goals with the outcome statements of the profiles; or it may simply use those outcomes that suit its curriculum. In either case, the outcomes form a framework by which children's achievements can be assessed and reported. The profiles approach to assessment and reporting relies on teacher judgement of children's achievement of specified outcomes. Because of this, curriculum-embedded tasks must give teachers oppor unities for observing children working on mathematical tasks, forming judgements about achievement and so contribute to the teacher's knowledge of children's relationships to the outcomes of the Profiles.

Using the Australian national profiles and its statements of learning outcomes together with a synthesis of curricula from several (Australian) state education systems, a collection of some two hundred tasks covering number, space, measurement, chance and data, and problem solving were created using the ground rules listed above. While not exhaustive with regard to any one state curriculum, the range of tasks offers a more

than adequate assessment resource for most elementary mathematics curricula. Activities were designed either for an individual child or for children working in pairs. In some cases activities were self-contained while others require a calculator, manipulatives or other extra equipment. Most activities were school-based, while a few involve work at home as well. Activities varied from closed to open-ended with a variety of response formats. These include paper-and-pencil, constructions, posters and written descriptions.

The sample activities below are examples of rough, first drafts. Each activity follows a recipe-like pattern. Activity 3.3 focuses on simple addition facts. The child being assessed is asked oral questions by a partner, who may be the teacher, and their response recorded. It was intended to make activities re-usable (administered to the same child more than once or that the two children involved could swap roles without loss of validity of the activity) so rather than a fixed set of questions, a random element has been introduced. A simple dropping of a button on the question grid selects the question and crossing-off tally boxes help keep track of the number of questions asked.

Activity 3.9 focuses on sketching simple 3-D shapes according to simple definitions or rules. Activity 3.11 has the student use manipulatives (buttons) for demonstrating basic fraction concepts. The responses are then recorded via drawings. Unlike the previous activities that are pencil-and-paper, in Activity 3.14 Multi-base Arithmetic Blocks (MAB) are used to 'make' numbers; these 'built' responses are then shown to the teacher for scoring.

After initial development activities were scrutinised by a panel of experts for both curriculum validity and test fairness. All activities developed were then piloted on a sample of students and teachers to ensure good face validity of the activities. At this stage scoring heuristics were developed and these too were piloted with a sample of teachers. Scoring of student responses on the activities was on a partial credit basis, that is, students were scored for partial success not just fully correct answers. Based on data from piloting, activities were refined, and final copies prepared. Each activity has a front child's page and a teacher's page on the reverse. Details of the scoring key and focus of the activity are to be found on the teacher's page.

The 'Gulliver' activity below is an example of a refined activity with its associated score key. Whilst only three scores are possible, each provides information about the child's ability to communicate a simple investigation. 'Cubes' on the other hand is assessing the child's ability to successfully complete an investigation, not communicate it and the scoring key illustrates this emphasis.

Calibration of all activities is to be carried out on an Australia-wide sample of students, and the analysis of trial data conducted using Quest© for partial credit responses. Calibrated activities will then be used to establish a developmental continuum for each aspect of mathematics (number, space etc). Student scores on the subset of activities selected for them places the student on this continuum, enabling teachers to assess both growth over time and a 'snap-shot' view of current performance. Descriptions of activities attempted are also placed on this continuum, providing teachers with immediate verbal reporting on student performance, providing easy reporting information for parents and children.

A calibration study means that in classroom use teachers would be able to select those activities which they deem to cover the curriculum for any individual (or group) and still provide a standardized assessment for them. It is not necessary for all children in a

You will need:

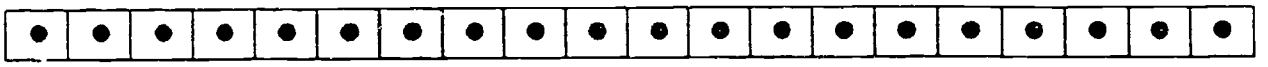
this page pencil button a partner

What to do

You will ask your partner 20 different questions.
To choose the question, drop the button on the chart below.
If you have already asked this question, drop the button again.
Read the question to your partner.

If your partner gets the right answer, put a ✓ in the question box.

Otherwise put a ✗.
Cross off a dot each time you ask a question.



When you have crossed off all the dots, write the number of right answers your partner got in the box at the bottom of the page.

2 + 4 + 2 = 8	6 + 4 + 3 = 13	6 + 7 + 2 = 15	8 + 8 + 3 = 19
6 + 2 + 1 = 9	3 + 7 + 4 = 14	9 + 3 + 2 = 14	4 + 9 + 4 = 17
1 + 1 + 7 = 9	9 + 1 + 1 = 11	6 + 6 + 1 = 13	6 + 9 + 3 = 18
7 + 1 + 2 = 10	4 + 6 + 4 = 14	8 + 3 + 3 = 14	7 + 7 + 5 = 19
2 + 2 + 2 = 6	2 + 8 + 2 = 12	7 + 5 + 3 = 15	3 + 8 + 7 = 18
4 + 3 + 2 = 9	8 + 2 + 1 = 11	8 + 0 + 6 = 14	7 + 9 + 2 = 18
4 + 0 + 5 = 9	5 + 5 + 3 = 13	5 + 6 + 3 = 14	6 + 6 + 4 = 16
3 + 2 + 3 = 8	1 + 9 + 2 = 12	9 + 2 + 4 = 15	2 + 9 + 5 = 16
1 + 2 + 2 = 5	7 + 3 + 0 = 10	8 + 4 + 1 = 13	5 + 5 + 9 = 19
4 + 1 + 4 = 9	10 + 0 + 5 = 15	6 + 7 + 2 = 15	4 + 7 + 7 = 18

NUMBER OF RIGHT ANSWERS: ☐

PARTNER'S NAME: ..... YEAR: .....

SCHOOL: .....

DRAW SHAPES

3.9

You will need:

this page

a pencil

What to do

Can you draw a box following these rules?

AND	SOME TRIANGLE SIDES	SOME SQUARE SIDES
SOME RECTANGLE SIDES		
SOME SQUARE SIDES		

NAME: _____ YEAR: _____

SCHOOL: _____

You will need:

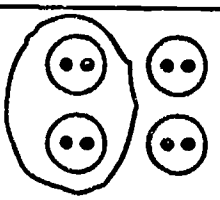
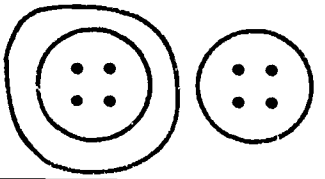
this page pencil buttons

What to do

Put out enough buttons to help you make these fractions. Make each fraction in two ways.

Draw the buttons you have used. Now draw a loop around enough buttons to show the fraction.

The first one has been done for you.

FRACTION	BUTTONS - one way	BUTTONS - another way
$\frac{1}{2}$		
$\frac{1}{3}$		
$\frac{1}{4}$		
$\frac{1}{5}$		
$\frac{1}{10}$		

NAME: _____ YEAR: _____

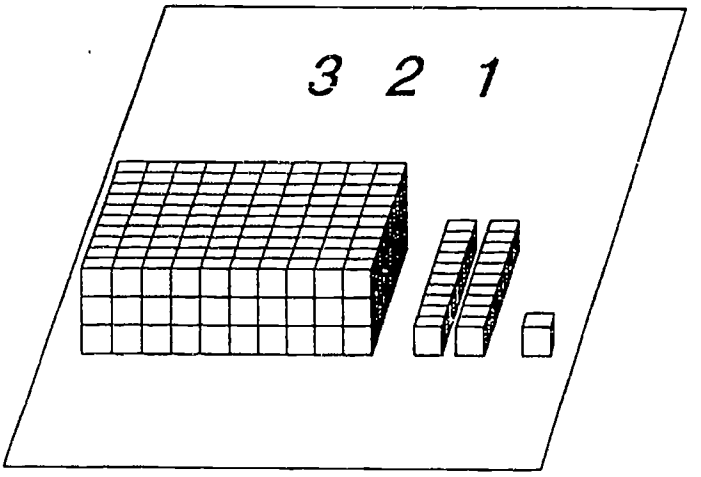
SCHOOL: _____

You will need:

this page MAB blocks pencil extra paper

What to do

Can you make numbers with the blocks?
This picture shows how to make the number 321.



Choose a number from this list.
Write it on a piece of paper.
Next to it make the number with MAB blocks.

121	302	535	720
916	784	212	604
405	263	981	342
540	830	152	863
353	674	483	671

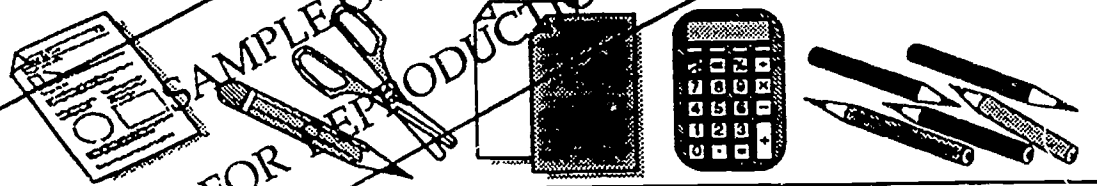
Choose more numbers until you have made ten numbers altogether.
Show your teacher what you have done.

NAME: YEAR:

SCHOOL:

GULLIVER

5-2



What to do

In the story *Gulliver's Travels*, Gulliver is supposed to be 12 times bigger than the people of Lilliput.

Imagine you are a Lilliputian.

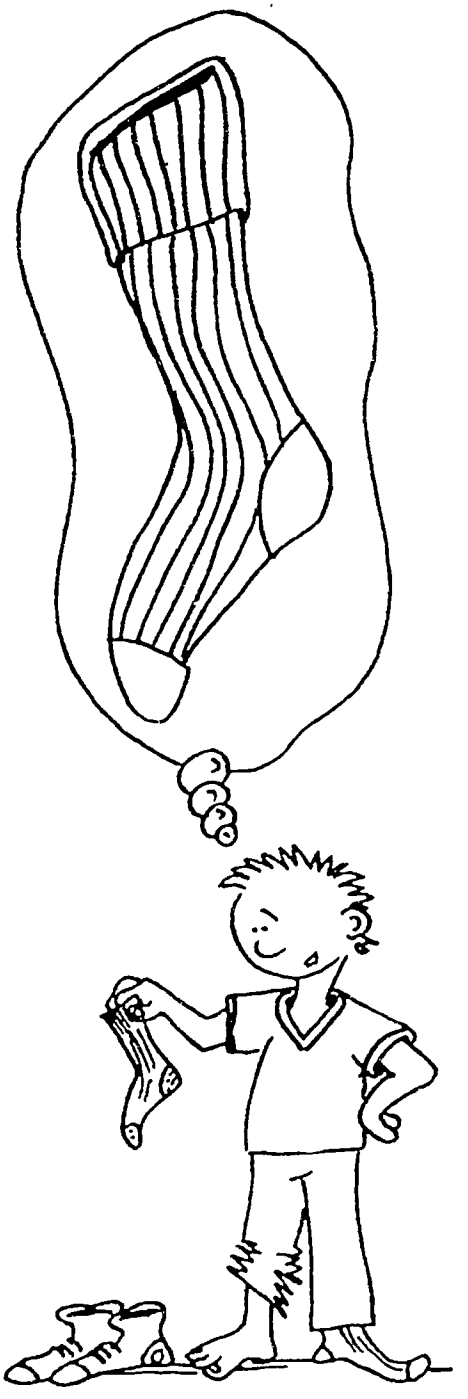
Measure your footprint and handprint.

Now use your calculator to find the size that Gulliver's footprint and handprint would be.

Make a display of your measurements and Gulliver's.

Mark the measurements on the display.

Write two or three sentences on your display to tell what you did and what you found out.



Show your work to your teacher.

What is being assessed

The child's ability to clearly communicate the results of a simple mathematical investigation.

Score	Description
2	The display has the dimensions marked. The explanation clearly describes the work done. (Clear enough for another to do the same.)
1	The display has the dimensions marked. The explanation does not make clear what has been done. (Not clear enough for another to do the same.)
0	Any other answer.

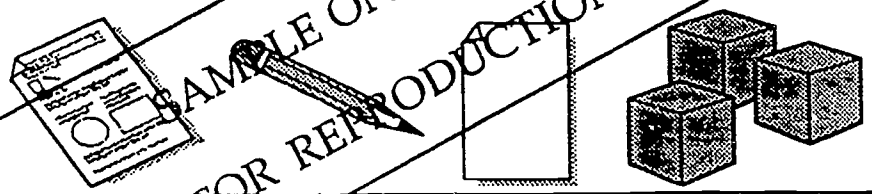
The maximum score for this activity is 2.

Comments

The scoring of this activity is based on the clarity of the explanation not on the accuracy of measurements.

CUBES

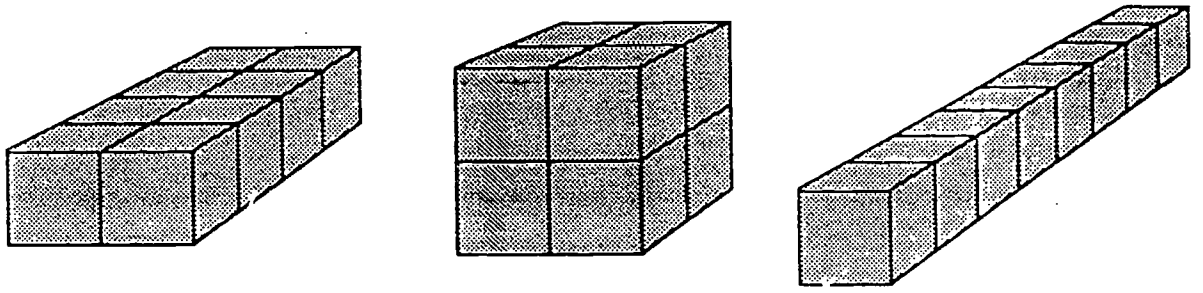
3-3



What to do

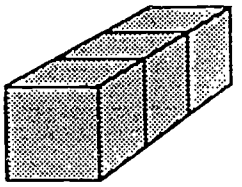
Using cubes you can make many box shapes.

For example, with eight cubes you can make:



Some numbers of cubes only make one box shape.

For example, with three cubes you can make:



Three is a one-box number.

Use your cubes to find out how many one-box numbers there are less than 20.

Make a list of these numbers.

Show your work to your teacher.

What is being assessed

The child's ability to carry out an investigation using concrete materials.

Score	Description	Answer
3	There is evidence that the cubes were used to find answers. The answer includes all nine primes less than 20 and no incorrect answers.	1, 2, 3, 5, 7, 11, 13, 17, 19
2	There is evidence that the cubes were used to find answers. The answer does not include the number 1 in the list or includes one incorrect answer; all other seven primes are listed.	
1	There is evidence that the cubes were used to find answers. The list has two or more omissions or two or more incorrect answers.	
0	The task could not be completed or there is no evidence that the cubes were used. Any other answers.	

The maximum score for this activity is 3.

Comments

As well as correct answers, these must be evidence that the cubes were used to complete this investigation.

class to do the same activities, but the reporting of achievement is still comparable across children. The aim of providing teacher choice and control seems to have been realized. Once suitable activities have been selected and administered, responses are scored and the raw score converted to a scaled continuum value. This is aligned on the continuum with verbal descriptions of activities whose difficulty lies in the same region of the continuum. This allows teachers to see at a glance the child's achievement, those activities which are easy for this child and those which are more difficult. Thus not only is assessment provided, but also some indication of activities suitable for future learning.

ACER is continuing to develop assessment material of the DART type and if the results of the calibration study show that this work is fruitful, we may see changes to the way we think about assessment. This material represents a breakthrough in standardized assessment practices as it allows teachers to integrate standardized assessment within their teaching, removing the necessity for off-the-shelf tests. Tailoring assessment to the needs of the child and the teacher is a first step towards beneficial assessment.

References

Australian Education Council. (1991). *A national statement on mathematics for Australian schools*. Carlton: Curriculum Corporation.

Doig, B. (1989). Rainy Day Activities. In *Blackline Masters Level Three, Young Australia Mathematics*. Melbourne: Thomas Nelson.