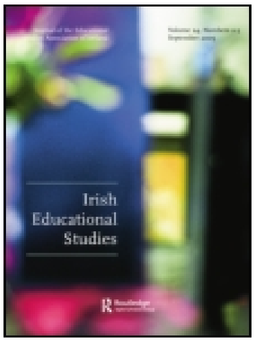


Irish Educational Studies

ISSN: 0332-3315 (Print) 1747-4965 (Online) Journal homepage: http://www.tandfonline.com/loi/ries20

Adapting science performance tasks developed in different countries for use in Irish primary schools

Paula Kilfeather , Michael O'Leary & Janet Varley

To cite this article: Paula Kilfeather , Michael O'Leary & Janet Varley (2006) Adapting science performance tasks developed in different countries for use in Irish primary schools, Irish Educational Studies, 25:01, 3-33, DOI: 10.1080/03323310600597501

To link to this article: http://dx.doi.org/10.1080/03323310600597501

Irish Educational Studies
Vol. 25, No. 1, March 2006, pp. 3–33

Adapting science performance tasks developed in different countries for use in Irish primary schools

Paula Kilfeather*, Michael O'Leary and Janet Varley
St. Patrick's College, Ireland

This article describes a four-year project undertaken to develop a set of performance tasks that could be used for assessing hands-on science in Irish primary schools. It begins by considering some of the literature on performance assessment and concludes with a discussion on the potential of the tasks to support teaching and learning in science. The main body of the article is structured to reflect the five phases of the research project itself. In phase one, science assessments used in a variety of educational systems in Australia, Canada, New Zealand, the United Kingdom and the United States were located and catalogued. In phase two, approximately 170 performance tasks were selected and adapted by the authors to suit the requirements of the Irish primary science curriculum. In phase three, a purposive convenience sample of teachers evaluated the extent to which the tasks (a subset of 67) were suitable for use at different grade levels. The teachers' feedback was used to amend tasks. In phase four, the researchers observed 11 different tasks being implemented in classrooms. The eleven teachers involved were interviewed about their experiences immediately afterwards. Again, based on the outcomes of this study, changes were made to the tasks. The fifth phase of the project, due to be completed in 2006, will involve the dissemination of 124 of the tasks to teachers via a booklet and a CD-ROM. Future prospects relating to other elements of the project such as Web-based resources, professional development courses and exemplars of performance are also discussed.

Background

Assessment in primary education may be described as the process of gathering, interpreting, recording, using and communicating information about children's achievements with respect to knowledge, concepts, skills and attitudes. The research described in this article was undertaken to examine the extent to which materials and methodologies used in countries with a history of science assessment at the primary level could be used to enhance and support the work of Irish teachers. The project was undertaken in response to the fact that a revised curriculum was introduced into Irish primary schools in 1999 and that, for the first time, science was named as a subject area for infants to sixth.[1] In addition, the principle of assessment as an

*Corresponding author. St. Patrick's College, Drumcondra, Dublin 9, Ireland. Email: paula.kilfeather@spd.dcu.ie

ISSN 0332-3315 (print)/ISSN 1747-4965 (online)/06/010003-31

DOI: 10.1080/03323310600597501

integral element of teaching and learning was espoused strongly in this new curriculum (Government of Ireland, 1999). Evidence from elsewhere suggested strongly that this type of assessment could be used as a tool for raising pupil achievement (Black & Wiliam, 1998).

Early on in the life of the project it became clear from the literature on classroom assessment and from discussions with individuals familiar with the different systems in each of the countries that many approaches and methodologies are used in the process of assessing science. These include teacher observation, questioning, conferencing, concept mapping, portfolios, projects, work samples, learning logs and commercial or teacher-designed tests and tasks (see, for example, Cowie & Bell, 1999; Harlen, 1999, 2000; Enger & Yager, 2001; Orpwood, 2001; Crooks, 2002). Many of these methodologies are discussed in the Irish primary curriculum documents and a number of them are likely to feature as topics when teachers receive in-service professional development on science over the coming years. While this project could have focused on all or some of these approaches—for example, how teachers in these countries go about observing their pupils, using work samples, organising portfolio assessment and the like—it seemed to us that principles of good practice were already available to teachers in the literature or had been made available to them during the curriculum days.[2] What was lacking in the Irish system but freely available in other countries was a plentiful supply of materials that had been developed to aid the process of assessing science through either traditional pencil-and-paper tests (mostly multiple-choice) or performance-based tasks.

In the literature, performance-based assessment is described as 'assessment which relies on the observation and judgement of activities as they occur' (Foster & Masters, 1996). Such assessments have three components: a task that requires pupils to solve a problem or to conduct an investigation using concrete materials in a hands-on way; a response format that allows pupils to communicate their findings; and a scoring system that allows judgements to be made about pupils' ability to carry out or complete the task. Brown and Shavelson (1996) point out that it is the inclusion of a *scoring system* that differentiates a performance assessment from a performance task.

Performance assessment is just one of an array of approaches that can be used in the classroom to gather information on pupil progress and achievement. Few would argue that it is the only type of assessment that should be used in the classroom or that it provides the most valuable kind of information about pupils. The reality is that different assessment approaches suit different purposes and research shows that performance-based assessment has not always led to the kind of learning gains predicted by its advocates (see, for example, Shepard *et al.*, 1996). There is general consensus in the assessment community that an array of approaches is required to build a full and accurate picture of a child's achievements and this is recognised in the Irish Primary School Curriculum (Government of Ireland, 1999). However, that said, performance-based assessment has a number of advantages that should be considered in the context of Irish classrooms. It is useful for gathering information on a wide range of achievements, including concept acquisition, communication skills, problem solving, critical thinking, psychomotor skills, and social/affective characteristics such as cooperation in groups. As Shepard (1997) argues, it is an approach that can be used by teachers to assess the extent to which pupils can transfer knowledge and skills to new situations (what Shepard calls 'robust understanding'). Performance assessment is also useful for charting progress over time, since performance assessment tasks can be used more than once with the same pupil without compromising their validity. Furthermore, it is useful for integrating assessment with teaching and learning (through the use of performance criteria), for identifying pupil learning needs, and for fostering pupil self-assessment (Linn *et al*., 1991; Linn, 1994; Baxter *et al*., 1996; Stiggins, 1997; Airasian, 2000; Kuhs *et al*., 2001). However, achievement gains are not an automatic result of introducing performance-based assessments. The issue is complex, with outcomes varying according to the cultural background of the learner (Bond, 1995), pupil ability (Arick *et al*., 1997), the curricular areas being assessed (Shepard *et al*., 1996; Fuchs *et al*., 1999) and the stakes involved for teacher and pupil (Koretz *et al*., 1994, 1996a,b). Another experience of performance assessments worth recounting is the large amount of time and effort required to construct, implement and score them (see, for example, Shavelson *et al*., 1992; Madaus & Kellaghan, 1993).

After much deliberation, an important decision was made that the research would focus on performance-based tasks rather than on, for example, paper-and-pencil multiple-choice or short answer tests. There were a number of reasons for this. First, the tasks seemed to provide an optimum approach for assessing skills and hands-on activities in a formative way. The performance-based approach made them especially compatible with the experimental and investigative nature of science, a key feature of the Revised Primary Curriculum for Science (Government of Ireland, 1999). The tasks also had the advantage of allowing for a balance to be struck between assessment, teaching and learning in a constructivist framework on the one hand, and a more scaffolded/teacher-directed one on the other. Second, the tasks seemed particularly suitable for assessing science *skills*, especially the skills associated with *Working Scientifically* and *Designing and Making*, as described in the 1999 Irish primary curriculum. Third, there were a large number of these tasks freely available for use in many countries around the world. A plentiful supply of performance-based tasks adapted for use in Irish primary schools was not available and, due to the enormous amount of work required in developing them, was unlikely to become available in the near future. Fourth, unlike the paper-and-pencil tests, the tasks had the potential to help teachers integrate assessment with teaching and learning. Fifth, the tasks had the potential to aid the assessment of higher-order thinking skills such as problem solving and reasoning. Finally, the tasks had the potential to provide formative, diagnostic and summative information about pupils.

Since the project began in 2001 it has evolved over five phases. The research methodology used and the findings with respect to each of the five phases are discussed in the following sections.

Mode of inquiry

Phase 1. Locating and cataloguing the tasks

Organisations and individuals involved in science assessment in five countries were contacted to identify extant materials used in primary schools. Tasks were sought from Australia, Canada, New Zealand, the United Kingdom and the United States (Doig *et al*., 1996; Harmon *et al*., 1997; EDWA, 1998; ASAP, 1999; NEMP, 2000; NZCER, 2000; QCA, 2000). The rationale for this choice was that these countries were English speaking and all had science curricula in place for primary schools. Permission was granted to adapt materials for use within an Irish primary context.

In all, we located over 1000 items measuring science achievement. These items were in different formats, mainly multiple-choice, short answer, essay and performance. About 500 were selected for inclusion in a database of items suitable for Irish primary schools. Following extensive cataloguing, attention was focused on approximately 200 performance-based tasks that sought to assess pupils' scientific concepts and skills. In that context it should be noted that a decision was made to concentrate on three of the strands only (*Living Things, Energy and Forces, Materials*). *Environmental Awareness and Care* was not included primarily because many of the concepts in this strand are attitudinal and are best assessed using approaches other than performance tasks.

Phase 2. Adapting the tasks

Materials were evaluated to establish how well they could be matched with the aims and objectives of the Revised Curriculum for Primary Science (1999) and aligned with the syllabus for each class level. Scoring procedures, assessment criteria and worksheets varied widely in the original sources and in some cases were absent. Considerable time was spent in adapting and creating suitable scoring rubrics and worksheets. A standard template or grid was devised so that assessment tasks appeared in a common format suitable for use in Irish primary schools (see Appendix 1 for an explanation of what information was included in each of the grid cells). In all, 170 tasks were adapted in this way. It was gratifying to note that we identified tasks suitable for assessing all of the process skills associated with *Working Scientifically* (questioning, observing, predicting, investigating and experimenting, estimating and measuring, analysing, recording and communicating) and also with *Designing and Making* (exploring, planning, making and evaluating).

A particular feature of the standard template was the creation of a rubric-style scoring guide that would facilitate teachers in making judgements about pupil achievement. The scoring guides were divided into four levels of performance (excellent, good, fair and weak) for both content and skill elements. While the scoring guides provided detail about what constituted performance at the four different levels, the intention was that teachers should assign a level for an individual performance on a 'best-fit' basis, rather than trying to apply too many specific rules.

It should be noted that the curriculum handbook suggests that, up to and including second class, simple investigations should be carried out where the problem, materials and method are suggested by the teacher. After that, teachers are encouraged to provide more open-ended activities for pupils. With respect to the latter, the authors had found, in working with primary teachers, that they could be reluctant to use open-ended materials due to their lack of confidence in teaching science. As a result we felt it would be appropriate to include a mixture of tasks—some that were relatively prescriptive and some that involved more open-ended investigations. The four adapted tasks in Appendix 2 are intended to reflect this range.

Phase 3. Evaluating the tasks with teachers

Sixty-seven tasks were selected for evaluation, drawn from all classes between junior infants and sixth class. The tasks covered the strands *Living Things*, *Energy and Forces* and *Materials*. It was decided that each of these tasks should be sent to at least three teachers. Sixteen schools were chosen as a convenience sample, which represented a range of types, including mixed, single-sex, primary, infant and senior schools. A standard evaluation form was used to obtain feedback on eight aspects of the tasks on an agree/disagree basis, with space for further comments provided.

In total, 183 evaluation sheets were returned from 64 teachers in 12 schools, representing a school response rate of 75%. Each teacher evaluated a maximum of three tasks, each task drawn from a different strand unit of the science curriculum. All 67 tasks were ultimately evaluated by at least one teacher. Thus, the evaluations for tasks from the strand units represented views from a variety of teachers in different settings. Teachers provided a range of comments about the tasks in the open-ended section of the evaluation sheet. These were analysed and common

Table 1. Themes from teacher feedback in phase three

Positive Feedback	Aspects for Development
• Tasks provide a challenge	• Some mismatch of tasks to class level
• Pupil enjoyment	• More specific information on prior knowledge needed
• Potential for integration	• Likely investigation outcomes requested
• Clear and simply laid out	• Scientific reasons for outcomes requested
• Sufficient information to field questions	• More subject knowledge background for teachers needed
• Directions easy to follow; foolproof	• Safety – reassurance needed
• Resource lists and teacher questions useful	• Steps could be numbered for clarity
• Child-friendly worksheets	• Sample blank tables could be included
• Option to draw responses seen positively	• Mixed response to amount of writing required
	• Concern about oral assessment (infants)

emerging themes identified. These are summarised in Table 1. A summary of teacher responses to the evaluation statements is also presented in Table 2. Responses to different tasks from each strand unit of the curriculum have been combined.

These results indicate a positive response to the tasks. It was significant that when teachers were critical of aspects of the tasks, it was often with respect to the fact that we had overestimated their scientific knowledge. This feedback was particularly useful in so far as these respondents appeared to be typical of the many primary teachers who do not have a strong background in science or science teaching (for example, see 'aspects for development' column in Table 1). To examine the feasibility of the assessment tasks more closely, it was decided to trial a selection of these in the classroom. This was undertaken in phase four and is reported below.

Phase 4. Trialling the tasks in the classroom

During the first term of the 2003/4 school year, 11 of the tasks were evaluated 'in action' in different classroom settings. Teachers agreed to be observed using a task with their class and to participate in a semi-structured interview. While the three schools involved in this phase were all designated disadvantaged schools in the Dublin area, the teachers who worked in them were similar to most Irish primary teachers in that they had no special expertise in science education. Feedback was obtained from teachers on pragmatic aspects of the tasks and assessments.

Eleven teachers spread across all classes from infants to sixth tried out the tasks with their pupils. A standard-format record sheet was provided to all teachers to enable them record assessment decisions about individual pupils or groups of pupils (see Appendix 3). Each teacher was observed as they implemented the tasks and interviewed immediately afterwards. This allowed us to triangulate our data. The observations were carried out using a semi-structured recording schedule involving three dimensions of task implementation and five time periods.[3] Interviews were semi-structured and recorded using a mini-disc.[4] Nine major findings arose after analysis of the observation notes and interview transcripts. These are summarised in Table 3.

Phase 5. Revising the tasks and dissemination

On the basis of the feedback received from teachers during phases three and four of the project, further amendments were made to the set of 170 tasks. The descriptions of the science concepts to be assessed and prior knowledge required to do the task were improved. The section on procedures was expanded to include more information for teachers on background science knowledge, about implementing each task and safety issues. A great deal of work was done to improve the scoring guides by including examples of likely student responses to illustrate different levels of understanding. These responses were derived from three sources: the responses of pupils to the tasks in the five countries where the tasks originated; the literature on common misconceptions in science; and the authors' own experiences of teaching science.

Table 2. Teacher responses to evaluation in phase three

	Strand Unit Combined % Agreeing								
Evaluation Statement	Myself n = 21	Plants & Animals n = 23	Light n = 16	Sound n = 23	Heat n = 6*	Mag & Elect n = 25	Forces n = 21	Props & Char n = 26	Mat & Change n = 22
The task is appropriate for Class Level indicated.	100	87	75	83	100	96	90	88	95
The concept being assessed is stated clearly.	100	100	100	100	100	96	95	92	86
The task focus is clear.	95	91	88	100	100	100	90	77	91
The description of the prior knowledge required is clear.	100	91	88	78	100	96	86	92	91
The skills being assessed are clearly identified.	90	100	100	87	100	96	95	88	95
Relevant information about the task is provided for the teacher.	90	91	94	83	100	96	81	81	86
Directions for how to do the task are clear.	95	91	81	87	83	100	95	81	91
The pupil worksheet is clear and relevant.	90	83	81	87	83	88	86	85	86

*Note: The number of teachers rating *Heat* tasks is low.

Table 3. Findings from observations and interviews in phase four

Outcome	Example
1. The tasks encouraged active involvement of pupils and teachers in the assessment process.	*Most of the time devoted to the task included pupils working together actively and teachers circulating amongst them.*
2. The teachers liked the tasks.	"*. . . the tasks don't come across as intimidating.*"
3. Teachers adapted them to suit their needs.	*Teacher did not make individual notes on the proffered record sheet . . . noted down general pointers for the whole class.*
4. The pupils liked the tasks; they *enjoyed* the assessment.	"*They really got into it. They were very enthusiastic about it and most of the talking was about the task*."
5. The tasks helped to integrate planning, teaching and assessment.	"*I would definitely use the scoring or assessment guide here. It will help me focus when I'm setting up experiments.*"
6. The tasks focused attention on individual pupil learning and needs.	"*[The tasks] . . . make you tune more into the children, make you aware of what . . . you are covering that individuals are not achieving*."
7. The teachers indicated that they would use the assessment information in different ways.	"*As a result of watching them do this . . . I will mix the groups in a different way.*"
8. Teachers were surprised by what they found.	"*Initially I thought the concept would be too easy for them but when I saw them doing it in practice, [it] was harder for them.*"
9. Teachers found that performance assessment is hard to do.	"*[I need] a lot more practice in doing that kind of wo*rk."

Note: Interview data are in quotation marks.

Spaces for the pupil's name and the date were added to the worksheets and, in some, tables were added to facilitate recording by pupils. The amount of writing required was reduced in a number of tasks by including the use of drawings as a substitute. Four examples of tasks and scoring guides, revised according to these criteria, are presented in Appendix 2. In addition, tasks that were deemed to be similar to each other, too difficult or easy, or unsuitable for the primary classroom were deleted from the set of tasks to leave a final set totalling 124 tasks. As Table 4 illustrates, we were reasonably successful in matching suitable assessment tasks with each of the curriculum strands and strand units at each of the class levels (the exception being the strand unit *Heat*).

Work has already begun on developing a CD-ROM of these tasks and an explanatory booklet for teachers. This will provide many teachers with a tangible

Table 4. Total set of tasks matched to the strands and strand units at each class level

Class LevelStrand/ Strand Unit	Infants	First/Second	Third/Fourth	Fifth/Sixth	Total
Living Things					
Myself/Human Life	3	2	5	5	15
Plants and Animals	5	6	5	3	19
Energy and Forces					
Light	1	5	1	2	9
Sound	3	2	4	1	10
Heat	1	0	1	4	6
Magnetism/Electricity	2	2	3	3	10
Forces	5	4	4	6	19
Materials					
Properties/Characteristics	2	5	4	8	19
Materials and Change	3	3	5	6	17
Total	25	29	32	38	124

resource that can be made available to them in their schools. Task descriptions, scoring guides and worksheets may be printed directly from the CD-ROM and photocopied. A website is also being planned. The intention here is not just to facilitate dissemination but to aid the process of updating the tasks over time. It may also be necessary to extend the project beyond the currently funded time frame to include work involving the development of exemplar materials (hard copy and video) showing pupil performance at different levels. In-service support could also serve to increase teachers' skill in using the tasks to integrate assessment with teaching and learning.

Conclusion

This research project has helped to identify concrete examples of suitable performance-based assessments that can be used in Irish primary schools to aid the process of science assessment. Through a process of adapting, evaluating and trialling, 124 tasks are now well aligned with the Irish primary science curriculum. These tasks are focused on key science concepts and skills and are performance based in that they require an active engagement with the scientific process. They require pupils to develop and use scientific concepts and skills to solve scientific problems. In that sense they are teaching and learning tasks as much as assessment tasks.

For each task we have tried to identify the criteria by which performance will be judged. While the rubric scoring system used is particularly useful for summative assessment, teachers can also observe the different parts of the performance suggested by the criteria used in the rubric for more formative or diagnostic purposes. These criteria can focus attention on what needs to be taught by the teacher and learned by the pupils. Indeed, the literature on classroom assessment would suggest that in most instances pupils should be made aware of the criteria for success before being assessed (see, for example, Stiggins, 1997). We feel that many of the tasks provide practical examples of key scientific skills in action and as such have the potential to be used as a resource during a lesson to improve the teaching process and to deepen the child's scientific understandings. While many of the tasks are more prescriptive at the lower class levels, others are extremely open-ended. Open-ended tasks feature more prominently at the upper class levels. This reflects the increasing pupil autonomy in decision-making that is a feature of skill development across the Irish primary science curriculum. In addition, our belief is that the tasks are flexible enough to be used not only as assessments, but also as resources for science learning when the scoring guide is not the primary focus for the teacher.

The evidence from the development phases of the project is that while the tasks present a number of assessment challenges, teachers can use them to improve teaching and learning in Irish primary school classrooms. Many of those who cooperated in our research indicated that the tasks provided them with opportunities to observe children's scientific skills in action, to identify strengths and weaknesses and to provide them with new insights into children's understanding of concepts. Our research also suggests that the tasks have the potential to provide teachers with an adaptable and practical way of integrating assessment with teaching and learning.

The extent to which our work in putting this resource together will bear fruit will depend on whether teachers and pupils find the tasks practical and useful. We set out to make the tasks accessible and comprehensive enough for teachers with relatively little experience of teaching practical science. Our hope is that the tasks will provide a support for teachers in their quest to become more comfortable teaching and assessing hands-on science. Another group who may have an interest in the tasks are the *cuiditheoirí*[5] who will be responsible for supporting the implementation of the science curriculum over the coming years. Our ability to review and update the tasks will also have a bearing on their long-term prospects. Needless to say, the provision of an adequate programme of in-service professional development dedicated not just to performance assessment but to assessment in general will also be crucial to their survival. In that context, it is clear that other instruments that measure scientific knowledge and children's attitudes to science should also be developed in time. Moreover, assessment approaches such as pupil conferences, concept mapping, curriculum profiling and the like should continue to be on the agenda of future teacher professional development initiatives for science and other curricular areas.

Acknowledgements

The authors gratefully acknowledge the help and support provided by participating primary schools, teachers and pupils in the evaluation and trialling phases of the project. This research was generously funded by a grant from the Research and Development Committee of the Department of Education and Science.

Notes

1. The previous curriculum contained a subject called Social and Environmental Studies. This consisted of Human Environment, Animal Life and Plant Life with suggestions for nature expeditions and investigation tables. It was only in the syllabus for fifth and sixth class that the term 'Elementary Science' was used.
2. During these days, teachers attended professional development courses related to the 1999 curriculum in their schools. Science was allocated two curriculum days.
3. The three dimensions were:
   - Task and pupil issues (prompts: Working alone/pairs/group? Discussion/questioning? Skills in evidence? Use of materials/worksheet? Enjoyment? On task?)
   - Task and teacher issues (prompts: Initial directions to pupils? Classroom management? Interaction with pupils? Explaining/narrating? Use of rubric/record sheet? Formative or summative assessment?)
   - Other (prompts: Prep done beforehand? Task appropriateness? Concept clarity? Suitability for which skills? Resources including worksheet? Time?)

   A new page for recording observations under these headings was used every ten minutes.
4. The questions asked were: What did you learn about individual pupils' skills and knowledge? Did the scoring guide help you to identify pupil achievements and challenges? What use will you make of this information? Did you use the record sheet? How? Was it useful? Manageable? (Or do you think you would you use the record sheet? How? etc). Was the task description sufficiently clear? Was the worksheet appropriate? (What would you change?) How do you feel your pupils reacted to the task? (What would you change?) Overall, do you think tasks like these would be useful for assessing hands-on science? What challenges do you think these tasks would present to teachers in general? Do you have any other comments to make?
5. 'Helpers' or subject experts.

Bibliography

Airasian, P. W. (2000) *Assessment in the classroom* (2nd edn) (New York, McGraw Hill).

Arick, J., Nave, G. & Jackson, P. (1997) *A full evaluation of the Oregon supported education plan and its impact upon student outcomes: final report*. ERIC Document: ED408794.

Assessment of Science and Technology Achievement Project (ASAP) (1999) *Classroom assessment in science and technology* (Ontario, Nelson Thomson Learning).

Baxter, G., Elder, A. D. & Glaser, R. (1996) *Assessment and instruction in the science classroom*. ERIC Document: ED406434.

Black, P. & Wiliam, D. (1998) Assessment and classroom learning, *Assessment in Education: Principles, Policy and Practice,* 5(1), 7–74.

Bond, L. (1995) Unintended consequences of performance assessment: issues of bias and fairness, *Educational Measurement Issues and Practice,* 14(4), 21–24.

Brown, J. H. & Shavelson, R. (1996) *Assessing hands-on science. A teacher's guide to performance assessment* (Thousand Oaks, CA, Corwin Press).

Cowie, B. & Bell, B. (1999) A model of formative assessment in science education, *Assessment in Education: Principles, Policy and Practice,* 6(1), 101–116.

Crooks, T. J. (2002) Educational assessment in New Zealand schools, *Assessment in Education: Principles, Policy and Practice,* 9(2), 237–252.

Doig, B., Cheeseman, J. & Lindsey, J. (Eds) (1996) *Exemplary assessment materials: science* (Victoria, Australia, Australian Council for Educational Research [ACER]).

Education Department of Western Australia (EDWA) (1998) *Monitoring standards in education—student achievement in science. Teacher's administration instructions and student's booklet: Energy* *and change; Natural and processed materials; Life and living* (Victoria, Australia, Australian Council for Educational Research [ACER]).

Enger, S. K. & Yager, R. E. (2001) *Assessing student understanding in science* (Thousand Oaks, CA, Corwin Press).

Foster, M. & Masters, G. (1996) *Performances: assessment resource kit* (Melbourne, Australia, Australian Council for Educational Research [ACER]).

Fuchs, L., Fuchs, D., Karns, K., Hamlett, C. L. & Katzaroff, M. (1999) Mathematics performance in the classroom: effects on teacher planning and student problem solving, *American Educational Research Journal,* 36(3), 609–646.

Government of Ireland (1999) *Primary school curriculum: science* (Dublin, The Stationery Office).

Harlen, W. (1999) *The teaching of science in primary schools* (London, David Fulton Publishers).

Harlen, W. (2000) *Teaching, learning and assessing science 5–12* (London, Paul Chapman Publishing).

Harmon, M., Smith, T. A., Martin, M. O., Kelly, D. L., Beaton, A. E., Mullis, I. V. S., Gonzalez, E. J. & Orpwood, G. (1997) *Performance assessment in IEA's third international mathematics and science study (TIMSS)* (Chestnut Hill, MA, Boston College).

Koretz, D., Stecher, B., Klein, S. & McCaffrey, D. (1994) The Vermont portfolio assessment program: findings and implications, *Educational Measurement: Issues and Practice,* 13(3), 5–16.

Koretz, D., Barron, S., Mitchell, K. & Stecher, B. (1996a) *The perceived effects of the Kentucky Instructional Results Information System (KIRIS), MR-792-PCT/FF* (Santa Monica, CA, RAND).

Koretz, D., Mitchell, K., Barron, S. & Keith, S. (1996b) *The perceived effects of the Maryland School Performance Assessment Program, CSE Technical Report No. 409* (Los Angeles, Center for the Study of Evaluation, University of California).

Kuhs, T. M., Johnson, R. L., Agrusa, S. A. & Monrad, D. M. (2001) *Put to the test* (Portsmouth, NH, Heinemann).

Linn, R. L. (1994) Performance assessment: policy, promises and technical measurement standards, *Educational Researcher,* 23(9), 4–14.

Linn, R. L., Baker, E. L. & Dunbar, S. B. (1991) Complex performance-based assessment: expectations and validation criteria, *Educational Researcher,* 20(8), 15–21.

Madaus, G. F. & Kellaghan, T. (1993) The British experience with authentic testing, *Phi Delta Kappan,* 74(6), 458–469.

McMillan, J. H. (2001) *Classroom assessment: principles and practice for effective instruction* (Boston, Allyn and Bacon).

National Education Monitoring Project (NEMP) (Dunedin, New Zealand). Available online at: http://nemp.otago.ac.nz (accessed March 2000).

New Zealand Council for Educational Research (NZCER). *Assessment resource banks*. Available online at: http://www.nzcer.org.nz (accessed March 2000).

Orpwood, G. (2001) The role of assessment in science curriculum reform, *Assessment in Education: Principles, Policy and Practice,* 8(2), 135–151.

Qualifications and Curriculum Authority (QCA) (2000) *Test Base* (London, QCA).

Shavelson, R. J., Baxter, G. P. & Pine, J. (1992) Performance assessment: political rhetoric and measurement reality, *Educational Researcher,* 21(4), 22–27.

Shepard, L., Flexer, R. R., Hiebert, E. H., Marion, S. F., Mayfield, V. & Weston, T. J. (1996) Effects of introducing performance assessment on student learning, *Educational Measurement: Issues and Practice,* 15(3), 7–18.

Shepard, L. A. (1997) *Measuring achievement: what does it mean to test for robust understanding?* (Princeton, NJ, Policy Information Center, Educational Testing Service).

Stiggins, R. J. (1997) *Student-centered classroom assessment* (2nd edn) (Upper Saddle River, NJ, Prentice Hall).

Appendix 1. Science assessment task template

<table>
<tr><td>Task:</td><td colspan="3">Each task was identified by a name that provides an indication of what the task involves, e.g. All afloat, Magic magnets, etc. In the published version of the grid this cell will be colour coded according to strand, e.g. red for Energy and Forces, etc.</td></tr>
<tr><td>Class:</td><td>Tasks were deemed to be suitable for a particular pair of adjacent classes as in the curriculum documents, e.g. first/second</td><td>Strand:

Strand Unit:</td><td>The task was matched to the strand and strand unit of the Irish primary curriculum, e.g. Strand Energy and Forces, Strand Unit Forces</td></tr>
<tr><td>Concept:</td><td>The scientific concept that underpins the task was identified, e.g. floating depends on a number of factors such as the shape of an object, etc.</td><td>Task Focus:</td><td>This was written to give the teacher a quick summary of what the task involved, e.g. pupils will sort objects into those that float and those that sink.</td></tr>
<tr><td>Prior Kn:</td><td>Here the teacher was provided with an indication of what pupils should have learned or experienced in order to attempt the task.</td><td>Skills Focus:</td><td>The skills being assessed were identified, e.g. predicting, investigating, observing, etc.</td></tr>
<tr><td>Materials:</td><td colspan="3">The section provided a full list of materials required to complete the task.</td></tr>
<tr><td colspan="4"><u>Procedure</u>
Three pieces of information were included in this section:
1. Information about the scientific knowledge, concepts and skills that might be useful for a teacher to know before undertaking the assessment with pupils;
2. Safety information;
3. A step-by-step guide on how to administer the task.</td></tr>
</table>

<table>
<tr><td></td><td>Strand/Strand Unit
Scoring Guide</td><td>Working Scientifically
or
Design and Making
Scoring Guide</td></tr>
<tr><td>Excellent</td><td colspan="2">Having conducted a literature review on performance-based assessment a decision was made to use rubric-style scoring guides as they were considered to be easier for teachers to use and would facilitate both formative (if teachers concentrated on a small number of the rubric elements) and summative assessments. In addition, as a result of reading the work of such science education experts as Harlen (1999) and McMillan (2001), rubrics were constructed for both the content (strand, strand unit) and process (working scientifically, designing and making) elements of each task. Again, using the literature as a guide, four levels of performance were described for most tasks. The terms ‘excellent’, ‘good’, ‘fair’ and ‘weak’ were chosen for the four levels as these terms were already in common use in schools (for marking, in summer reports and the like). Examples of pupil responses at different levels were provided for many tasks.</td></tr>
<tr><td>Good</td><td></td><td></td></tr>
<tr><td>Fair</td><td></td><td></td></tr>
<tr><td>Weak</td><td></td><td></td></tr>
</table>

Appendix 2. Examples of adapted science assessment tasks

Task:	**Too loud!**		
Class:	Infants	**Strand:** **Strand Unit:**	Energy and forces Sound
Concept:	Some materials make loud or soft sounds when shaken.	**Task Focus:**	Pupils will make shakers and investigate loud and soft sounds.
Prior Kn:	Pupils should be able to identify loud and soft sounds and use appropriate words to describe materials, e.g. hard, soft.	**Skills Focus:**	Exploring/investigating Making Evaluating Recording
Materials:	Margarine tubs with lids, or plastic cups with lids. Blocks or Lego bricks, cotton wool, paper clips, grass, tissue (or items such as these). Worksheet (optional).		

Procedure

Teacher information.
Pupils may need help putting covers on the containers.
Encourage children just to try one material at a time.

How to do this task

- See if you can make two shakers that make **loud** sounds.

1. Which of these materials do you think would you use?

- Use the materials one at a time to make the shakers.

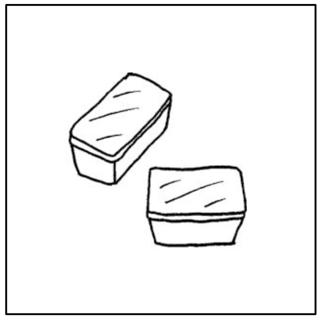
2. Do the shakers make a loud sound?

- Look at the pictures. Tick the things you used to make loud shakers.

Too loud!

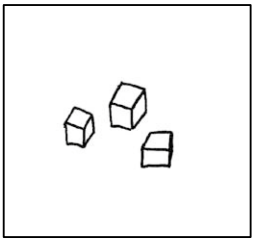
	Strand/Strand Unit Scoring Guide	Designing and Making Scoring Guide
Excellent	Demonstrates a clear understanding of the concept. Both shakers constructed make loud sounds. Pupil accurately identifies the blocks and paper clips as suitable materials.	Selects all appropriate materials. Constructs both shakers, changing contents if necessary to make loud shakers. Tests the shakers. Records accurately.
Good	Demonstrates some understanding of the concept. One shaker makes a loud sound. Pupil correctly identifies blocks **or** paper clips as the material to make the loud shaker.	Selects materials (at least one of which is not appropriate) and constructs two shakers. Tests the shakers. Some recording accurate.
Fair	Demonstrates little understanding of the concept. One shaker makes a loud sound but is unsure about which material was used for this shaker.	Constructs at least one shaker. Some aspects of testing and recording are accurate. Pupil may test a range of materials together.
Weak	Demonstrates no understanding of the concept. Cannot distinguish between loud and soft sound and/or does not make a shaker with a loud sound. Unable to identify materials that make loud or soft sounds.	Does not select appropriate materials. Appropriate shakers are not constructed. Purpose of activity not understood.

Name: ____________________ **Date:** _________

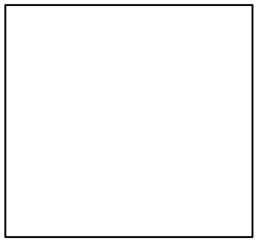


Too Loud!

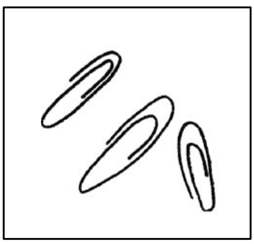
✓ the things you used
to make **loud** shakers.



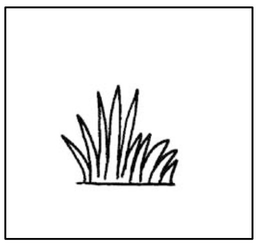
blocks



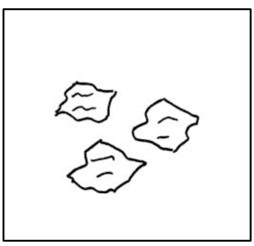
cotton wool



paper clips



grass



tissue

Task:	**Wind Car**		
Class:	First/Second	**Strand:** **Strand Unit:**	Energy and Forces Forces
Concepts:	Air can be used to move things. The material used in a sail and its size and shape can affect how well an attached object moves.	**Task Focus:**	Pupils will design and make a vehicle that is powered by wind.
Prior Kn:	Pupils should be able to identify objects that are moved by wind, e.g. windmill, yacht. Pupils should be able to label drawings.	**Skills Focus:**	Exploring Planning Making Evaluating.
Materials:	Lego block with wheels or wheeled toy. Material for sail, e.g. dowel, lollypop stick, J-cloth, paper towel, thread, string, card, cardboard, paper, tin foil, scissors, plasticine, Sellotape, Blutak. Worksheet.		

Procedure

Teacher information.
Purpose of this activity is to explore types of sail that work rather than to construct wheels that work. Therefore, Lego block with wheels attached (or wheeled toy) is a good starting point.

How to do this task

- Your challenge is to make a wind car that works well.
- Look at the picture on the worksheet.

1. How do you think the wind car moves? (Oral question).

- Look at the sail.

2. What do you think would make a good sail for your model? (Oral question).

- Allow time for pupils to handle the materials and plan how they are going to make their wind car.
- Draw the model that you made. Use the words on the worksheet to help you to label your drawing.
- Show me how your model works

3. Are you happy with the way it is working? Why? Why not?
4. Can you think of ways to make the sail work even better?

Wind Car

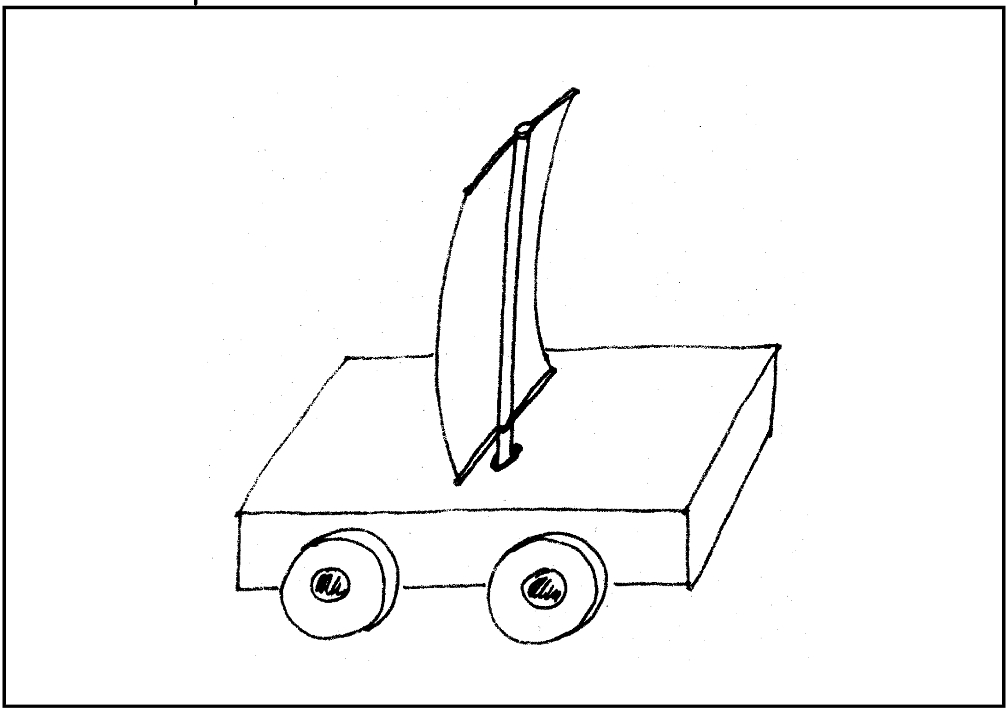
	Strand/Strand Unit Scoring Guide	Designing and Making Scoring Guide
Excellent	Demonstrates a clear understanding of the concepts. In Q1 and in testing his or her own model the pupil identifies air/wind as the cause of movement. In Q2 and 4 the pupil identifies more than one factor for making an effective sail, i.e. size, shape, type of material.	Demonstrates a thorough understanding of the problem to be solved. Selects all appropriate materials and adapts them where necessary. Tests the product and records the results accurately. Evaluates the model and makes suggestions for modifications (Q3). The diagram/drawing is very clear and all key labels are included . The product fully addresses the original task set.
Good	Demonstrates some understanding of the concepts. In Q1 and in testing his or her own model the pupil identifies air/wind as the cause of movement. In Q2 and 4 the pupil identifies one factor for making an effective sail.	Demonstrates a basic understanding of the problem to be solved. Selects some appropriate materials and adapts them where necessary. Tests the product and records results. Evaluates the model and makes one suggestion for modification (Q3). The diagram/drawing is reasonably clear, e.g. some key labels missing. The product adequately addresses the original task set.
Fair	Demonstrates little understanding of the concepts. Plausible response to either Q1, 2 or 4. May suggest blowing harder in response to Q4 (inappropriate response as it does not refer to sail)	Demonstrates a partial understanding of the problem to be solved. Selects some appropriate materials but makes few adaptations. Tests the product but does not record the results. Suggestions for modifications are irrelevant e.g. I would paint it (Q3). The product partially addresses the original task set.
Weak	Demonstrates no understanding of the concept. Responses to Q1, 2, and 4 are either missing, incorrect or inappropriate (e.g. Q4: I'd colour it in).	Demonstrates little understanding of the problem to be solved. Selects few appropriate materials and makes no adaptations to them. May test product inappropriately e.g. pushes it. The product does not address the original task set.

Name: ________________ Date: _____________

Wind car

Your challenge is to make a wind car that works well.

Here is a picture of a wind car.



To talk about:

1. How do you think the wind car moves?

   - Look at the sail.

2. What do you think would make a good sail for your model?

Draw the model that you made. Use the words below to help you to label your drawing.

Helpful words:

Lego wheels paper marla card string sail stick cloth

To talk about:

3. Are you happy with the way it is working?
4. Can you think of ways to make the sail work even better?

Task:	**Mighty Magnets**		
Class:	Third/Fourth	**Strand:** **Strand Unit:**	Energy and Forces Magnetism and Electricity
Concepts:	Some magnets are stronger than others. Strength of magnets can be measured.	**Task Focus:**	Pupils will find out about the strength of three magnets. They will design their own way of measuring magnet strength.
Prior Kn:	Pupils should have used different magnets, e.g. for magnetic attraction and sorting activities. Pupils should have undertaken other investigations using a fair test.	**Prior Skills:**	Observing Investigating (fair testing) Recording Estimating and measuring
Materials:	Per group: 3 different magnets, labelled A, B and C or clearly different in appearance. Paper clips (these could be of different sizes for an extra challenge) One cent coins Pieces of thick card Ruler Tray or container for equipment Worksheet.		

Procedure

Teacher Information

Ensure that magnets provided to each group are of different strengths. You may need to test this beforehand. There are a number of different ways in which a valid strength test can be done and it is important that pupils try to come up with their own ideas.

Acceptable ideas which the pupils might have include:

Counting number of clips picked up in a clump;

Counting number of clips picked up, chain fashion (one paper clip attracts the next);

Counting the number of coins/clips you can drag along the table in a chain;

Measuring how near you can bring a magnet to a paper clip before it 'jumps' towards the magnet;

Counting the number of pieces of card needed to block the effect of the magnet.

Some pupils may struggle to select a sensible test. For these pupils, reduce the available test materials, and prompt with suggestions only if they are still struggling after this.

Safety

Pupils should keep magnets away from computer disks and screens.

How to do this task.

- Pupils are to make up a test to find out which magnet is strongest, using any of the things on the desk.
- Allow time for pupils to explore all the materials before they decide on a test that will work. Only give pupils one magnet during this planning phase.
- Write and draw to show how you will test your magnets.

1. How will you make sure your test is fair? Write down your ideas.

- Give the pupils the other two magnets to test.
- Now test your magnets.
- Write or draw to show what happened.

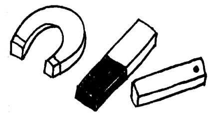
2. Which was the strongest magnet?
3. How do you know this magnet was the strongest?

Mighty Magnets

	Strand/Strand Unit Scoring Guide	Working Scientifically Scoring Guide
Excellent	Demonstrates a clear understanding of the concepts. Pupil identifies one magnet as being the strongest (Q2). This is clearly based on the measurements taken, e.g. this magnet picked up 15 clips, so it is the strongest (Q3).	Clear and efficient procedures for conducting a fair test of the strength of the magnets are identified and followed. Relevant data are gathered and analysed. Information is recorded accurately and is very well organised and communicated. The conclusion reached about which magnet is strongest is consistent with the data gathered and is accurate, e.g. The bar magnet picked up 15 clips. This was the most, so it is the strongest.
Good	Demonstrates some understanding of the concepts. Pupil identifies one magnet as being the strongest (Q2). This is based on the data collected but explanation does not refer to exact measurements taken (Q3), e.g. this one picked up the most clips.	Procedures for conducting a fair test of the strength of the magnets are developed but are somewhat limited in their efficiency (e.g. some inconsistent use of materials or measurements). Relevant data are gathered and analysed. Information recorded is mostly accurate and is reasonably well organised and communicated. The conclusion is consistent with the data gathered.
Fair	Demonstrates little understanding of the concepts. For Q2 pupil may identify incorrect magnet. For Q3 suggests that the biggest magnet was the strongest. Pupil shows reluctance to accept the evidence of their testing e.g. pupil persists in retesting largest/ horseshoe magnet because they are convinced it must be the strongest.	Procedures for conducting a fair test are developed but are very limited in their efficiency (e.g. pupils may only realise when questioned afterwards that their tests were not fair). Some relevant data are gathered and partly analysed. Information recorded may be inaccurate and/or poorly organised and communicated. The conclusion does not make reference to the data e.g. I know the horseshoe magnet was strongest because it is the biggest.
Weak	Demonstrates no understanding of the concepts. May identify a magnet in Q2 (may be a guess). Offers no plausible explanation for Q3.	None of the components for a fair test is identified. No observable set of procedures for conducting a fair test is followed. No recognition that tests were unfair. Little or no data are gathered, recorded or communicated. The conclusion is lacking, inaccurate or based on guesswork. Pupil is unable to offer reasons for any conclusion drawn.

Name: ___________________ Date: ___________

Mighty Magnets



Your challenge is to find a way to test magnets to find out which one is strongest.

What you can use:

3 magnets, paper clips, 1 cent coins, pieces of thick card ruler

Plan how you will carry out your strength test.
Write and draw to show what you will do to test your magnets.

1. How will you make sure your test is fair?

Write your ideas here.

____

____

____

____

Now test your magnets.
Write or draw to show what happened.

2. Which was the strongest magnet? ________________

3. How do you know this magnet was the strongest?

____

____

Task :	**Absorbency**		
Class:	Fifth/Sixth	**Strand:** **Strand Unit:**	Materials Properties and Characteristics
Concept:	Materials are chosen for their properties, e.g. some types of paper are more absorbent than others and are therefore better for mopping up water.	**Task Focus:**	Pupils will design a fair test that will measure absorbency in different kinds of paper.
Prior Kn:	Pupils should be familiar with the term 'absorbent'. Pupils should have experience of carrying out investigations involving fair testing.	**Skills Focus:**	Observing Predicting Investigating (fair testing) Recording Estimating and measuring Evaluating
Materials:	Per group: Different kinds of paper (e.g. different paper towels, tissue paper), water, tray, transparent jars or containers, rubber bands, scissors, droppers, plastic straws, stop watch or timer, container that measures volume in ml., ruler. Worksheet.		

Procedure

Teacher Information

All the papers available for this test must be absorbent, so that the outcome is not obvious from the outset. The final conclusion may vary between groups, depending on the type of investigation carried out.
There are several ways in which an effective test can be carried out – encourage pupils to explore their ideas first. It is not necessary for all groups to conduct exactly the same test. Only provide suggestions if pupil is struggling.
Examples of suitable tests might include: time taken for a liquid to reach a certain point when moving up the paper; measuring volume of liquid soaked up in a specific time (by measuring amount of liquid in container before and after soaking); jar could be covered by paper and they could count the number of drops of liquid added before it starts dripping through the paper.

Safety

Care needs to be taken to avoid water spillage on floor.

How to do this task

- Allow time for pupils to read worksheet.
- A paper company is trying to decide which paper is the most absorbent for soaking up spills.
- Each group is to design a test to choose the best paper for the company. The testing method should be clear to others if they wish to check the results themselves.
- List the equipment you need to do the test. (Allow time for pupils to explore and briefly experiment with the equipment before finalising their test strategy. Only give out one type of paper at this point so that pupils focus on designing a suitable test.)
- Draw and write to explain what you will do to test the papers.

1. How will you make sure this is a fair test?

- Now select four types of paper to test.

2. Which paper do you think will be the best? Why?

- Now carry out your tests.
- Record your results in the table (pupils may need help).

3. Which paper would you recommend to the company?
4. Why?
5. If you could do this investigation again, what would you do to make your test better? Explain why.

Absorbency

	Strand/Strand Unit Scoring Guide	Working Scientifically Scoring Guide
Excellent	Demonstrates a clear understanding of the concept. Response to Q.2 indicates a relevant feature of the predicted material e.g. it has holes in; it is thick; it is not shiny. Response to Q.4 suggests either that paper soaked up the most water or that paper absorbed water the fastest.	Demonstrates a thorough understanding of the purpose and design of the test. In Q.1, pupil refers to at least two variables that need to be kept the same e.g. same size paper; same amount of water; same length of time. Clear and efficient procedures for conducting a fair test are followed. Relevant data are gathered, analysed and explained. Information is recorded accurately and is very well organised and communicated. In Q.3, pupil's choice of paper is the one that has the best absorbency according to their data. Q.5 makes at least two suggestions for improvements and explains them e.g. I would make the pieces of paper bigger because it would be easier to see the difference.
Good	Demonstrates some understanding of the concept. Response to Q.2 suggests a feature of the predicted material although it may not be entirely relevant e.g. it is soft. Response to Q.4 does not compare with other papers, e.g. it soaked up a lot of water; it absorbed water quickly.	Demonstrates a basic understanding of the purpose and design of the test. In Q.1, pupil refers to one variable that needs to be kept the same. Procedures for conducting a fair test are followed but are somewhat limited in their efficiency. Relevant data are gathered, partly analysed and explained. Information is mostly accurate and is reasonably well organised and communicated. In Q.3, pupil's choice of paper has the best absorbency according to their data. Q.5 makes one suggestion for improvement and explains it.
Fair	Demonstrates little understanding of the concept. Response to Q.2 is either circular or does not refer to any observable properties e.g. it is the best; it's like the paper in the TV ad. Response to Q.4 does not refer to properties, e.g. it's the best; it will soak up the most water.	Demonstrates a partial understanding of the purpose and/or design of the test. In Q.1, pupil's suggestions for fair testing are vague or inappropriate, e.g. I will work carefully; we will take turns. Procedures for conducting the test are developed but are very limited in their clarity and efficiency. Some relevant data are gathered and partly analysed and explained. Information recorded is somewhat inaccurate and poorly organised and communicated. In Q.3, pupil's choice of paper has good absorbency according to their data, but is not the best. Pupil may re-state predicted material, even though data does not support it as the best. Q.5 makes one suggestion for improvement, but this is vague and is not explained e.g. I'd take longer to do the test.
Weak	Demonstrates no understanding of the concept. Responses to Q.2 and Q.4 are either absent or irrelevant e.g. it has a nice pattern.	Demonstrates little understanding of the purpose or design of the test. Response to Q.1 is either absent or irrelevant, e.g. I will wipe up the spilled water. No set of procedures for conducting a test is identified. Testing is done in a haphazard manner. Little or no data are gathered, recorded or communicated. The response to Q.3 is absent or not supported by the data. Response to Q.5 is either absent or irrelevant.

Name: ________________ Date: __________

Absorbency

A paper company is trying to decide which is the most absorbent paper for soaking up spills.

Your group has been asked to design a test to choose the right paper for the company. Your testing method should be clear to other if they wish to check the results themselves.

Your teacher will show you the equipment you can choose.

- List the equipment you will need for your test.

- Draw and write to explain what you will do to test the papers.

1. How will you make sure this is a **fair** test?

___

___

___

- Now select four types of paper to test.

2. Which paper do you think will be the best?

___

Why?

___

___

- Now carry out your tests.

- Record your results in this table.

Paper type	Results/Observations

3. Which paper would you recommend to the company?

____

4. Why?

____

____

5. If you could do this investigation again, what would you do to make your test better? Explain why.

____

____

____

Appendix 3. Record sheet used by teachers during the trialling of tasks in the classroom with example records

Date: *26-11-2003*	**Class:** *Third*		**Performance Task Name:** *Pulse rate**	
Pupil Name	**Concept Rating**	**Skills Rating**	**Strengths**	**Areas for Improvement**
John Allen	Ex	Ex		
Jane Baker	G	G	*Very careful labelling of diagrams 17/11*	*Didn't evaluate initial results 3/10 Evaluation improved 17/11*
Mary Cooke	F	G	*Systematic planning and recording 3/10*	
Alan Dell	G	W		
Etc				

*Note: Strand/Strand Unit =Living Things/Myself