IAEA Conference 2008

Soft Skills - Hard to Assess?

An investigation into enhancing the reliability of practical assessment tasks

Presented by Barbara Hill

Abstract

Since 2002 a number of initiatives in Scottish education have led to an increased emphasis on soft or core skills which have often been considered difficult to assess reliably. This paper will describe some attempts to increase the reliability of practical assessment tasks in the context of the Scottish Survey of Achievement (SSA).

The SSA was introduced in 2005 and is an annual sample survey. Its introduction was intended to separate national monitoring from classroom-based assessment and lessen the perceived negative wash-back effect on teaching and learning. The data gathered allows the Scottish Government to track standards of attainment at a national level – and at local authority level for those authorities who choose to opt for this.

The principle aim of the SSA is to produce national estimates of achievement for pupils across Scotland at different stages in their education. However, another objective is for the survey to play a role in demonstrating effective and innovative assessment approaches. The survey includes classroom-based tests, questionnaires and practical assessment tasks including Working with Others and Problem Solving. These are delivered and assessed by visiting 'field officers' (teachers trained to carry out the assessments).

This presentation will look at these innovative assessment tasks with a view to seeing whether soft skills are assessable in a valid and reliable way. This presentation will explore some of the different types of practical assessment

activities used in the survey and focus on the ways in which we attempt to maximise inter-rater reliability.

Introduction

This paper, and the presentation, will attempt to show that, in educating young people for the ever-changing workplace of the future, soft skills will become ever more important and, in turn, the need to assess these skills in a valid and reliable way will become key. Soft skills are generally regarded as difficult to assess or measure but, as they are increasingly valued, the ability to assess them in a reliable and robust manner is imperative.

I will endeavour to put this report into the context of the Scottish national assessment system, to describe the work that has been undertaken so far, and to outline my plans for future research in this area. This paper will describe the first stage of an on-going study into the reliability of the assessment of Problem Solving and Working with Others tasks which are administered as part of the SSA.

Background and context
The Scottish Survey of Achievement (SSA)

The publication, in 2004, of the results of a national consultation on assessment, testing, and reporting ('Assessment, Testing and Reporting 3-14: Our Response') showed clearly that what was wanted, by the teaching profession and the general public, was a well-balanced assessment system that put due emphasis on supporting children's learning, as well as providing accurate information about national standards of achievement. There was a commitment in the report to introduce a sample-based survey, the SSA, which would provide an overview of attainment levels at national level. It was envisaged that this would provide '…a robust national monitoring system that provides accurate information about overall standards and trends in achievement…' without distorting classroom practice. (http://www.scotland.gov.uk/Publications/2004/11/20177/45857)

In addition, a secondary aim of the SSA was to demonstrate innovative or different methods of assessment by providing exemplification of alternative methods to traditional paper and pencil tests.

The SSA covers four main curriculum areas: English Language, Social Subjects, Science, and Maths in a rolling programme with a different curriculum area being the focus of the survey each year. All of the pupils involved in the survey complete two written booklets and a sub set of the pupils also take part in a practical component.

Soft skills and the practical component of the SSA

In the Scottish curriculum the skills of Communication, Problem Solving, Working with Others, Information Technology, and Numeracy are described as *core* skills whilst in England they are referred to as *key* skills. Increasingly, however, there is a tendency to describe personal/interpersonal skills (problem solving, working with others, and communication) as *soft* skills and, for the purpose of this paper, that is the term I will use. I have already touched on the growing importance of these skills

Core skills are assessed as part of the SSA each year. By including assessment of these skills as part of a national survey, the SSA recognises and promotes the importance of these skills - *'There has been broad agreement that all … people need a set of personal attributes and skills that will prepare them for both employment and further learning.'* (Balgobi, Hutton, Rees, Tank Consulting Ltd., Weinstock, 2004, p8) In order to assess skills such as Problem Solving and Working with Others in an authentic manner they are included in the practical component of the survey and assessed by visiting 'field officers'.

Following each survey, the results are published in a brief 'Headline Report' and in a more detailed 'Supporting Evidence' report. The results of the practical component always carry a 'health warning' highlighting the perceived unreliability of this aspect of the survey:
'*Field officer judgements about pupils' abilities are necessarily subjective and, although training was provided to help improve the consistency of their*

assessments, judgements were not subject to moderation. The results should be treated as indicative.'
(http://www.scotland.gov.uk/Publications/2004/11/20177/45857)

The project
The main aim of the project was to explore ways in which the reliability and validity of the practical component might be enhanced. A review of the literature related to this topic might best be summed up in the words of LR Gay – '*Establishing reliability is a prerequisite for establishing validity.*' (Gay L R, Selection of Measurement Instruments, 1987)

This report and presentation focus on one of a number of areas I intend to investigate and describes my efforts to improve the consistency of inter-rater reliability through improving the training of the field officers. Inter-rater reliability has been defined as *'...the concern that a student's score may vary from rater to rater.*' (Moskal, Barbara M & Leydens, Jon A) These variations between markers or assessors can, and do, occur in any type of assessment, but even more so in the assessment of processes and soft skills such as problem solving and working with others.

Investigating the effects of different types of rater training
The SSA practical tasks are assessed in a variety of ways, including through group work and individual tasks, by 'field officers' - practising teachers who have been nominated by their local authority to be seconded to the project. The field officers undertake training and then work in pairs, visiting between five and six schools to carry out the assessments.

Prior to 2007 field officer training consisted of a small amount of task exemplification material being sent out to participants prior to their attending a face to face training day before visiting the schools. In 2008 it was decided to expand and enhance the training by developing a blended course – an online component and a face to face training day. The online component, SQA Academy (http://www.sqaacademy.com/moodle/), was designed to allow the teachers to develop familiarity with the practical tasks through using video clips

4

for assessment and discussion and, therefore, promote a shared understanding of the standard.

In addition to the practical assessment issues covered in the online component, a short theory unit was introduced – an overview of the principles and practice of assessment.

In order to gather data which might indicate whether the new training was having the desired effect, it was decided very early in the training course that participants would be asked to assess a task and to submit their assessment judgements online. The field officers were asked to watch a clip of four pupils working together to solve the problem illustrated in Figure 1 and to use the assessment criteria assess the pupils.
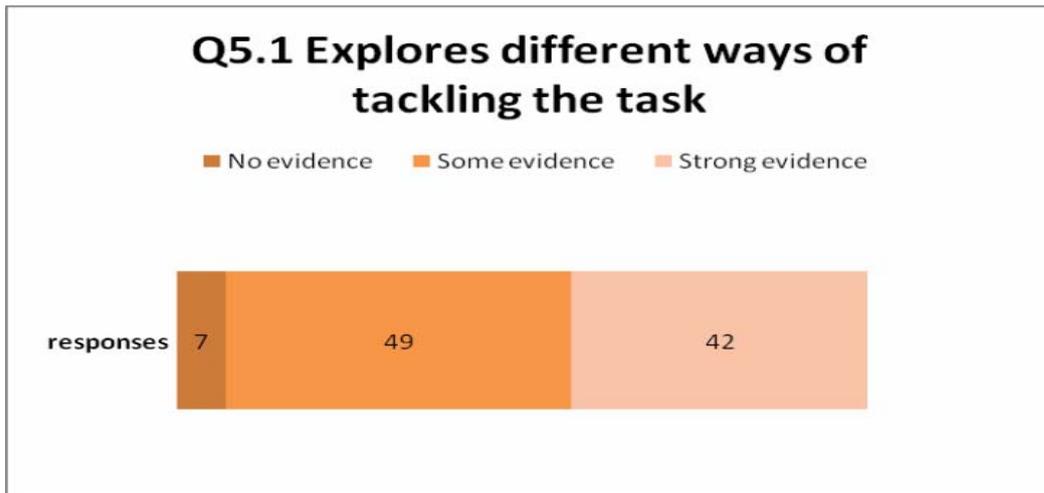
Figure 1: Extracts from the group problem solving task and the associated assessment criteria

Working with Others/Problem Solving

Working with Others/Problem Solving
Task 9b
 	
Who Gave What?
 	
YOUR CHALLENGE:
Eight classmates decide to hold a sponsored walk to raise money for their school trip later in the term.
 	
   o   Each pupil walks as fast as they can round the route, starting from school and finishing back at school.
   o   Each pupil makes a contribution to the school trip fund
 	
Working with your group:
   •   Use the clues to fill in a card for each child, showing
      o   the order they arrived (1st, 2nd, 3rd and so on);
      o   how much money they contributed.

5.Planning *Planning &* *organising*	1.Explores different ways of tackling the task
	2.Comes to a consensus as to strategy, course of action
	3.Completes task

From the table below it can be seen that of the 98 participants who submitted their assessment of this activity only 7 felt that there was no evidence of the group exploring ways of tackling the task. The remaining ninety one field officers were fairly evenly split (42/49) in deciding whether there was *some* or *strong* evidence of the group exploring different ways of tackling the task. It was this lack of agreement that we hoped to address through the training.

Figure 2: Assessment judgements early in training

Following each assessment clip field officers were encouraged to discuss the video clips and other assessment issues in the discussion forums. The quality of these discussions was very high, quickly leading to much reflective discussion. In turn this activity appeared to lead to a shared understanding of the assessment standard and an emerging assessment literacy amongst the group. Extracts from these discussions can be seen in figure 4.

Figure 4: Extracts from the SQA Academy forum discussion

'I found it difficult to assess whether the boy in red identified strengths and weaknesses of his own contributions.'

'*I also found it difficult to assess whether the group identified strengths and weaknesses as I felt they focussed on strengths.'*

'... felt lacked in evaluation ... apart from the one girl being a bit quiet, they worked extremely well in the group.'

'I quite like these forums as I find typing out my response helps me to shape up my thinking.'
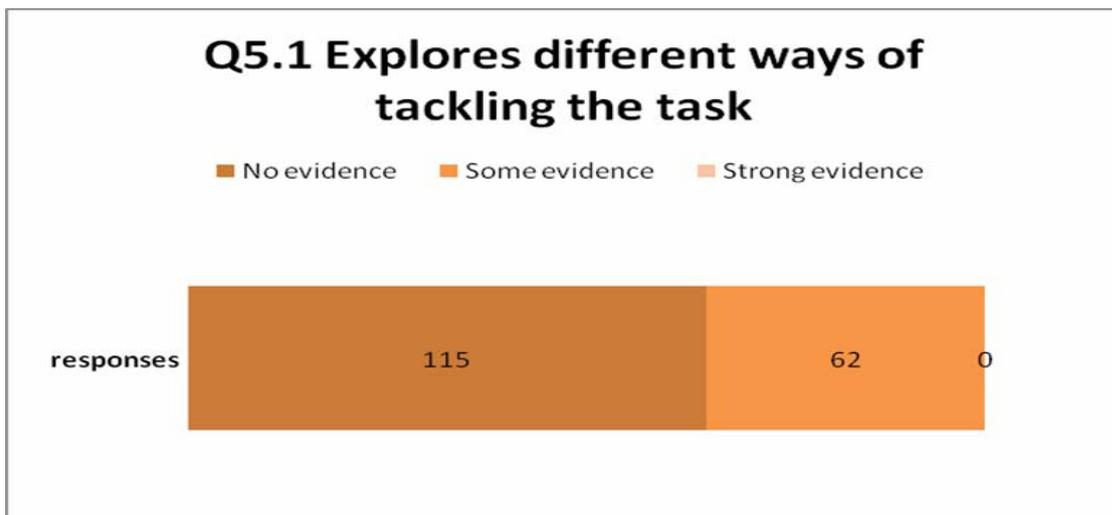
> *'It is useful to be able to see these videos and even having now seen only two I am feeling more confident about the process.'*
>
> *'... agree with earlier comments that threads on discussion forums are also invaluable as they often stimulate greater thought and added depth to task.'*
>
> http://www.sqaacademy.com/moodle/

The data gathering exercise was repeated on the face to face training days using different video clips but the same tasks. On these days the field officers watched the video clips and made their assessment judgements using an interactive voting system. This time, as can be seen below, the responses were split between *no* and *some* evidence but this time a significant majority agreed there was no evidence of the group exploring different ways of tackling the task.

Figure 5:Assessment judgements from training days

Conclusions

This data appears to show that the level of agreement between field officers had indeed increased to some extent following training. This pattern of greater agreement was seen to some degree across all the tasks assessed on the training days and may show that the extended/improved training does increase inter rater agreement in the SSA practical assessments. The new style of training certainly encourages participants to become familiar with the tasks and to develop confidence in assessing areas that are traditionally considered

7

difficult to assess. In addition, it became apparent that in allowing an opportunity to discuss assessment and assessment issues – both online and on the training days - was greatly valued by participants and led to an understanding of a shared standard. Another important by-product of the new style of training was the development of a network of assessment literate teachers across Scotland.

What next?

The investigation described here is just the beginning and I intend to continue this study and to investigate other factors that may impact on the reliability of this important type of assessment. This will include: trying to identify the characteristics and attributes of expert raters; exploring different ways of moderating the assessment judgements made by field officers; and a review of the practical task assessment criteria, because well-designed clear criteria can reduce the occurrence of discrepancies or differences in assessment judgements.

The introduction in Scotland of *Curriculum for Excellence*, which aims to prepare every child and young person from 3 to 18 for life in the 21st century, means that the type of assessment tasks described here, the evaluation of critical thinking and of the ability to transfer and apply these skills, will become even more important. It is, therefore, important that the assessment tasks demonstrated through the SSA are reliable because '*An assessment system cannot be valid without being reliable*' (Testing and Assessment, House of Commons Report 7 May 2008) and the ability to assess soft skills in a reliable and valid manner is essential in ensuring that the SSA remains fit for purpose and supports the Scottish national assessment system.

References

Gay L R, (1987) *Educational Research: Competencies for Analysis and Application*

Gay L R, (1987) *Selection of Measurement Instruments*

House of Commons Report, (7 May 2008) *Testing and Assessment*

Mertler, Craig A, (2001) *Designing Scoring Rubrics for Your Classroom*

Moskal, Barbara M, (2003) *Recommendations for Developing Classroom Performance Assessments and Scoring Rubrics*

Moskal, Barbara M, (2000) *Scoring Rubrics: What, When and How?*

Moskal, Barbara M & Jon A Leydens, (2000) *Scoring Rubric Development: Validity and Reliability*

Payne, Jonathan (2000) *The unbearable lightness of skill: the changing meaning of skill in UK policy discourses and some implications for education and training*, Journal of Education Policy, 15:3, 353 — 369

Scottish Government Report, (June 2008) *Scottish Survey of Achievement: 2007 Science, Science Literacy and Core Skills – Supporting Evidence Report*

Simpson, Simon, (2006) *The Measurement and Recognition of Soft Skills, Developing a Common Standard?*

SQA report *(*2003) *Reviewing the Core Skills Profile*

U.S. Department of Education (March 2006) *Results That Matter, 21st Century Skills and High School Reform*