

Assessing the creativity of scientific explanations in elementary science: an insider-outsider view of intuitive assessment in the hypothesis space*

Douglas P. Newton, Durham University, UK

Abstract

Assessing creativity is commonly believed to be difficult but there is evidence that an intuitive, holistic assessment is easy and reliable. Given that children can engage in creative activity and teachers are expected to foster it, some assessment of it could aid planning and optimise support. Assessing creativity intuitively and holistically could be a useful, quick way of assessing to inform teaching. A teacher of young children, however, is essentially an outsider in the child's world and often also in the world of science. Judging a child's creativity in this way from a child's point of view may not always be easy or reliable. Here, pre-service elementary teachers in the UK assessed explanations of simple science events. Their holistic assessments of creativity did not agree to any great extent. Agreement improved when they assessed some of the attributes of creativity but was still less than expected. Nevertheless, these novices' assessments as a whole showed there was some ability to discriminate usefully between explanations, albeit with considerable variation from teacher to teacher. Some implications for teacher training are described.

* A version of this appears as: **Newton, D.P.** (2010) Assessing the creativity of scientific explanations in elementary science: an insider-outsider view of intuitive assessment in the hypothesis space, *Research in Science & Technological Education*, 28(3), 187-201. ISSN 0950-0693 & 1464-5289 DOI 10.1080/09500690903233249

Introduction

Creativity is often highly valued (Craft, 2005; Kaufman and Sternberg, 2006). In the West, this is partly because its products can have economic value but also because, amongst other things, it can enhance personal autonomy (Newton & Newton, 2009a). Schools have been seen as stifling creativity but teachers are now encouraged to foster it amongst their students (e.g. DfES, 2003). The assessment of creative work in science could inform planning, optimise support and help to personalise learning in the classroom. But can it be assessed?

Creativity

Creativity is commonly taken to mean successful activity intent on producing something novel. Without intent, the product is only an accident and no merit accrues to the person; without novelty, nothing has been created, only reproduced. Intent and novelty, however, are not enough, the product must also be validated by society which rules on what is appropriate, suitable, effective or valuable (Csikszentmihalyi, 1996). But, while the creative act might be satisfying, enjoyable, even exhilarating, these are not necessary attributes: ‘Creativity needs mental discipline, prior experience and knowledge.’ It is not simply fun and free expression (Gouge & Yates, 2002, p. 137). The essential dimensions of creativity – intention, novelty and value – are its defining features in more recent definitions, such as that of NACCCE (1999, item 29): ‘Imaginative activity fashioned so as to produce outcomes that are both original and of value’. Additional attributes seen as worthy are elegance (Besemer & O’Quin, 1987, 1999), ethical acceptability (Cropley, 2001) and wisdom (Craft, 2008). Elegance describes a well-crafted, aesthetically pleasing, simple, concise or economical product. Wisdom refers to a consideration of how, or if, that product should be used.

Creativity is not the preserve of a gifted few: we can all be more or less creative as, for instance, when we solve everyday problems (Amabile, 1983a, Boden, 2004). Scholars like Ausubel (1978) and Csikszentmihalyi (1996), however, have described children as generally incapable of creative thought as they are unlikely to produce something both new to the world and of value. Others, like Williams (1970) and Petty (1997) argue that children can produce something which is at least new to them. This ‘new to the person’ view has been called psychological or ‘little c’ creativity (Richards, 1993; Gardner, 1993; Boden, 1996). The ‘new to the world’ view has been called historical creativity or the ‘big C’ view.

Children being creative in science

Science aims to describe, order and explain the natural and physical world (Jardine, 2000). From a constructivist perspective, generating explanations and testing them are creative processes (Newton, 2000). Klahr and Dunbar (1998) describe constructing explanations as working in ‘the hypothesis space’ and constructing tests of those ideas as working in ‘the experiment space’. In the elementary science classroom, children could be given a wooden metre rule and told to drop it so one end strikes the floor and bounces. The problem is: Why does it bounce? Thinking in the hypothesis space, these children construct various scientific explanations that are more or less new to them and, in the process, show psychological creativity. In practice, this kind of event can be used to build a ‘concept cartoon’, a pictorial representation with notes as an aide memoire to support thought in the experiment space (Keogh & Naylor, 1999). In the experiment space, children could be asked to devise a practical investigation which tests one or more of their possible explanations. Such constructions involve imaginative

thought, the bringing together of ideas and the generation of possible worlds. To the extent that their ideas are novel in their world, this is again an instance of psychological or small c creative activity. Some science curricula (for instance, Rose, 2009) include another mode of thought which could be called 'the application space'. This is where scientific knowledge and understanding are used to solve practical problems. Here, children might be asked to apply their understanding of the 'springiness' of a wooden lath to invent a device to close a door. The generation of likely explanations of the world, however, is central to the scientific endeavour (Jardine, 2000) although it tends to receive little attention (Roberts, 2004). This study focuses on assessing creative thought of this kind, namely, scientific explanations constructed in the hypothesis space.

The emphasis on the various essential and desirable qualities of creativity varies with context. Artists, for instance, tend to emphasise originality more than appropriateness, while architects emphasise function - their version of appropriateness - more than novelty. Generally, the creative work of architects is more strongly constrained by the brief, the situation and physics than that of artists so creating within those boundaries is central to their work (Glück, Ernst & Unger, 2002). Strongly constrained creative activity is often described as problem solving and 'imaginative problem solving is at the root of all human inventiveness' (Jardine, 2000, p. 5). That problem solving in science is a creative process was not generally acknowledged until the twentieth century. Before that, creativity was commonly seen as the domain of the poet and the artist, a view which still permeates popular notions of creativity (Tatarkiewicz, 1980; Euster, 1987; Treffinger et al., 2002) and those of some researchers who refer to the 'the creative disciplines' (e.g. de la Harpe & Peterson, 2008).

Assessing creativity

Creativity is both a process and a product. What a child does to generate an explanation, the explanation itself, or both, may be assessed. Which of these should be assessed has been the subject of debate for some time. For example, Fox (1963) and Rust (2002), referring specifically to creativity in science, argue that there is nothing unique about the creative process and attention should be on attributes of the product. Others (e.g. Treffinger & Poggio, 1972; Houtz & Krug, 1995) argue for assessing the process. In practice, there can be a strong correlation between assessments of the creative process and assessments of the creative product (Hennessey, 1994, p. 193) and examination of the product may offer clues about the processes which produced them. Accordingly, the product, the scientific explanation, is at the focus of attention in this study. It is potentially available for undivided attention after the event and assessments of it may guide teaching. Furthermore, the success of what the teacher does next is, at least in part, measured by the subsequent product (see e.g. Amabile, 1983a; Balchin, 2005) and that product, in turn, may offer further clues about what to do next. But how might children's scientific explanations be assessed?

Some teachers do not believe assessing creativity is possible because creative products are essentially unpredictable so cannot be anticipated with grading criteria (Rogers & Fasciato, 2005; de la Harpe & Peterson, 2008; Newton & Newton, 2009a). But assessment may be possible without grading criteria. Amabile (1983a, 1983b, 1996) has described 'consensual assessment' in which experts rate a created product intuitively according to what a creative product means to them. The consensus amongst the experts constitutes the outcome of the assessment. Cropley (2001, p. 101) concluded that assessing creativity

intuitively – without conscious reasoning (Colman, 2003) and largely using the ‘undermind’ (Claxton, 1997) – is not difficult as ‘intelligent observers’ make judgements which agree well with those of others. For example, Hennessey (1994) studied teachers’ intuitive, holistic assessment of elementary school children’s storytelling and found a high level of agreement (inter-judge reliabilities – a measure of agreement – ranged from 0.83 to 0.91). Similarly, Hickey (2001) used the approach to assess the musical compositions of 9 and 10 year-olds by various groups of teachers, including ‘general/choral music teachers’. Some 7 and 12 year-old children (13 and 24, respectively) also judged the compositions. The inter-judge reliability for these teachers was 0.81. The young children seemed to confuse liking and creativity and responded unreliably. This suggests that pre-service teachers’ could have some ability to assess creativity intuitively in the elementary science classroom.

There are, however, difficulties with this. The first is what I will call the insider-outsider problem. Cropley’s and Amabile’s confidence largely applies to insider assessment: adults familiar with the subject judge the creativity of other adults (Hennessey, 1994). But teachers are outsiders in the children’s world. Judging children’s creativity requires them to see products from a child’s point of view. For someone with limited knowledge of the norms of children’s thought, this is likely to be challenging. Primary teachers are also often outsiders as far as science is concerned. Many do not have a strong science background and their conceptions of scientific creativity can be narrow, even unsound (Newton & Newton, 2009a). For instance, the construction of a causal scientific explanation is not always seen as a personal, creative act while reproductive manufacturing activity following instructions may be described as creative (Newton & Newton, 2009b). Another obstacle is one of practicality. Consensual assessment calls for

several judges. In order to assess the artistic products of USA secondary school students, Beattie (2000) included secondary students amongst the expert judges with some success. The only true insiders in this world are the children but, unfortunately, their judgement can be distorted by cognitive limitations, such as an inability to differentiate between 'liking' and 'creative' (Hickey, 2001). Teachers of young children will often have to rely on themselves.

Aims

The insider-outsider model suggests that intuitive assessment will not be successful in all contexts. Assessing explanations holistically in the hypothesis space may be one of them. In the USA, Besemer and O'Quin (1987; 1989) helped untrained assessors with rating scales for various attributes of creativity (corresponding to novelty, logical plausibility/usefulness and elegance). With these scales, the judges achieved a high degree of agreement. This approach may be useful in this context. There is, however, more to assessment than agreement on the scores. There also needs to be some degree of validity, that is, judgements about creativity need some acceptable basis.

Accordingly, the questions addressed were:

- 1. To what extent are intuitive, holistic assessments of science classroom explanations reliable (where reliability refers to agreement between independent assessors)?**
- 2. Does assessing attributes of creativity intuitively increase the reliability?**
- 3. Does assessing scientific explanations in these ways produce valid results?**

Answers to these questions will inform discussion about the assessment of creativity in the hypothesis space.

Method

These questions were addressed empirically by having pre-service teachers at the end of their training course score sets of elementary science-classroom explanations and comment on the task. The first task was to assess creativity globally. Its purpose was to gauge the extent to which these novice teachers agreed in their holistic, intuitive assessment (Q1 above). The second task was to assess components of creativity (novelty, plausibility and elegance). Its purpose was to indicate the extent to which an intuitive assessment of components of creativity enhanced agreement (Q2). The quantitative and qualitative data from both tasks were used to inform discussion about the validity of the assessments (Q3).

Materials

Five assessment events were constructed, one for each of Sound, Ecology, Electricity, Dissolving and Forces. Each offered five, diverse ‘explanations’ capable of usefully informing the discussion. They were judged by three, primary science specialists to be explanations of the kind young children give in such contexts. The events spread over the UK’s Key Stage 1 (KS1: 5 to 7 years), lower Key Stage 2 (LKS2: 8 to 9 years) and upper Key Stage 2 (UKS2: 10 to 11 years; see Appendix). Age labels were attached to the events as these have been found to encourage judges to recognise the limitations of the children concerned (Hennessey, 1994).

Each event and its explanations had two forms. The first form asked for the explanations to be scored holistically and intuitively for scientific creativity on a scale of 0 to 5. The second provided a scoring grid with rows for the explanations and columns for novelty, scientific plausibility and elegance. There were also scoring rules which nested elegant in

plausible and plausible in novel explanations: all explanations had to be scored first for novelty, if zero was awarded an explanation for novelty, it was to receive zero for plausibility; if it was awarded zero for plausibility, it was also to be awarded zero for elegance. These *sine qua non* rules were to weigh against reproductive and implausible explanations. Each row was also summed to give a score out of 15 for each explanation.

Procedure

Participants read the instructions on the first form and were reminded that the task was to assess the explanations for creativity, relying on intuition. They were told to treat all explanations as intentional and without moral or ethical concerns. They were reminded that what counted was what was creative to the children and the explanations had to be judged from the child's point of view. After scoring, the participants had to indicate the explanation they believed to be 'correct', if they thought there was one.

The first form was turned face down to reveal the second form which participants read. Participants were told they were to assess some attributes of creativity intuitively. They were reminded that what counts here is what is novel, plausible and elegant for the children. Children may offer a variety of explanations but not all may be new to or original for the child. Furthermore, some could be more plausible than others. Natural, rational, mechanistic explanations are generally more plausible in science than supernatural and highly speculative explanations which do not relate well to the children's body of scientific knowledge (Berg, 2009). Of these, a child might see an explanation as more elegant, pleasing or economical than another (Besemer & O'Quin, 1999).

Together, the assessments took about fifteen minutes. Collaboration was not permitted. On completion, the participants were invited to comment on the task. Attention was focused on perceived task difficulty and its causes and on reasons for the ratings of the explanations.

Participants

Each event was assessed by groups of twelve pre-service, trainee teachers. Ecology, Dissolving and Forces were assessed by undergraduate trainees who had almost completed the final year of a three year degree course which would qualify them to teach in primary schools. Sound and Dissolving were assessed by postgraduate trainees who had almost completed their one year course which similarly prepared them to teach in primary schools. **The undergraduates were all female and had attended modules on science and elementary science pedagogy in each year of their course. During this time, they also were placed in primary schools for a total of 26 weeks where they taught a range of subjects, including science. One of the two postgraduate groups included two men. Two people (women) had science-related degrees. All had recently completed a taught course on elementary science teaching and had practised teaching in primary schools for a total of 22 weeks. The composition and educational experience of the participants were similar to that of the UK primary school teaching force (DES, 2005, 2006). All participants were volunteers and were free to withdraw at any time although none did so. No significant differences between the participants and the remainder of their cohorts were noted.**

Some two months before the rating exercise, all participants attended a one hour session on the meaning of creativity in the context of children learning elementary science. This

presented the view that all children can be more or less creative and illustrated opportunities for scientific creativity in the hypothesis, experiment and application spaces. No reference was made to assessing creativity.

Analysis

The scores are not on an interval scale (double the score does not indicate twice as much creativity) but are ordinal (that is, they are sufficient to place the explanations in a rank order from most to least creative). Coefficients of concordance provided measures of agreement between these rankings (Siegel, 1956). Collective and individual responses to particular kinds of explanation were collated and related to indicate the validity of the assessments.

Results

How well these pre-service teachers agreed in their intuitive assessment of the explanations is shown in Table 1. When scientific creativity was assessed holistically, the agreement between the teachers was statistically significant but, nevertheless, was generally low. When attributes of creativity – novelty (N), plausibility (P) and quality (Q) – were assessed individually, the teachers' agreement was often better. How these attributes may be combined to represent creativity scores as a whole is not known. Simply summing and ranking the scores, however, resulted in greater teacher agreement (S) than in the holistic assessment.

<< Table 1 >>

The rank order of the holistic creativity, novelty, plausibility and quality scores tended to be different indicating that the attributes were not treated as synonymous with the global concept. Nevertheless, the average rank order of S was usually close to that of the holistic assessment (Table 2).

<<Table 2>>

Some responses to particular explanations are now presented to inform discussion about the validity of the assessments.

1. Responses to 'correct' explanations

The explanations of each event included at least one which approximated to the accepted explanation, albeit in a simple, even simplistic, way. For example, in the Dissolving event, this was *b: Bits came off and broke up smaller and smaller....* Most pre-service teachers marked this as the 'correct' explanation and, as in the other events, considered it to be highly creative. Insofar as it is the child's own construction, this is not unreasonable: there is novelty for the child and scientific plausibility. But one person marked *c* as correct: *The water soaked into the sugar cube and now they're the same.* This was rated as the most creative both holistically and by components. In the Forces event, both *d* and *e* could be described as correct but *d: I've seen it in a book in the library ...*, was a reproduction while *e: If you look at it, the ruler bends ...*, could be a child's construction. Most felt that *e* was the correct explanation and considered it to be the most creative in both approaches. Two who marked *d* as correct also scored this reproductive explanation as the most creative both holistically and by components. (Those who applied the scoring rules when assessing components of creativity, in effect, discounted *d*.) The attraction of what was believed to

be the 'correct' explanation was generally strong. At least one was aware of this and commented that, 'It's easier in Art because there isn't a right answer.' Another reflected on the mental tension she felt in assessing the explanations of the Electrical Circuits event. She knew that *e: You need to hold the battery higher so the electric can run down to the bulb*, was not the accepted scientific explanation but felt that, 'The idea of lifting the battery higher has something about it. It's clever. It could be right for them.'

2. Responses to non-creative and irrelevant explanations

Some events also had 'explanations' which were reproductive or did not address the problem. Some participants noted this, as when one commented, 'It has been learned and not constructed creatively by the learner' and another said, 'It's just recall'. Already mentioned in the Forces event was *d: I've seen it in a book in the library ...* Being reproductive, it was expected to be appear low in the rankings but, in practice, it was highly regarded by some, perhaps because it was also seen as a 'correct' answer. In the holistic assessment, for instance, its rank ranged from 1 to 4. Likewise, *c: The ruler is made of wood. I bet we could make a ball that bounced*, does not attempt to explain the event and, again, should be discounted. One participant commented that, 'The child is not really answering why'. Nevertheless, in the holistic assessment, ranks ranged from 1 to 4; in the assessment by components the range was 2½ to 5. Similarly, in the Dissolving event *d: Too much sugar clogs things up*, seems largely irrelevant as an explanation of the event. Its ranking in the holistic assessment ranged from 1 to 5 while in the components assessment, it ranged from 3 to 5. In other words, while the overall tendency was to mark these explanations down, some rated them fairly highly, especially in the holistic assessment.

3. Responses to 'mediocre' explanations

Between these extremes, it would be reasonable to expect less plausible explanations, highly speculative explanations, weak analogies, and those of low novelty to appear. For example, in the Sound event, *d: It opens up your ears* has little basis in fact - even a young child is likely to question its veracity - so it is likely to have a low level plausibility. On average, it was never highly rated, although some did see creativity in it. Its rank ranged from 1 to 4 in the holistic assessment and 1 to 5 in the others. In the Forces event, *b: The Earth is like a big rubber ball*, appeals to a weak analogy, perhaps questionable even for a child. It was not highly rated. Nevertheless, it could be described as novel and it appeared third in the assessment of the novelty component. As with other explanations, the spread of individual rankings was relatively large (2 to 5). One participant described explanations like: *The battery is warming up* as not being as 'complex' as some of the others and she referred to its low ability to explain in a reasonable way. Given that children are often familiar with battery behaviour in toys, they may also find this explanation doubtful. It was rarely rated highly. Rating such explanations was not found easy. One participant commented: 'It is hard to say some explanations are more creative than others - they're so, sort of, equal.' This is probably a fair comment given that the various attributes of creativity may vary together but in different directions.

Discussion

The first question asked about the level of agreement when assessing the scientific creativity of explanations of events holistically and intuitively. Amabile (1983a, 1996) and Cropley (2001) expressed confidence in a shared, intuitive ability to assess creative products holistically with a high degree of agreement but the context was largely that of subject experts assessing older students' products. The data in Table 1 (Holistic column)

indicates that it would be unwise to extend this confidence to the context of pre-service, primary teachers' assessment of elementary scientific explanations. The agreement about the rank order of the explanations, reflected in coefficients of concordance ranging from 0.23 to 0.47 (mean 0.32), was generally low.

Besemer and O'Quin (1987, 1999) found teachers agreed more when assessing particular attributes of creativity. **The second question asked about the extent to which this was true of assessments of scientific explanations.** Table 1 (N, P, Q columns) provides some support for this with coefficients of concordance ranging from 0.25 to 0.80 (mean 0.49). Comparing both approaches, the average rankings produced by the holistic assessment tended to be very similar to those produced by summing the attributes (S in Table 1). For example, they were identical for the Ruler event (*e, a, c, d, b*) and very similar for the Ecology event (*d, a, c, e, b* and *d, a, e, c, b*, respectively). In other words, the effect of assessing attributes of creativity and summing the outcomes generally reduced the variability between the pre-service teachers (concordances ranging from 0.38 to 0.62, mean 0.53) but did not alter the rank order greatly. Summing attributes, however, needs caution. Here, these assessments were simply added to produce an overall creativity ranking (S). In reality, we do not know how or if these attributes should be combined to produce an overall assessment of creativity. Furthermore, in this study, there were some differences in the ratings of the attributes but if the intuitive assessment of one of the attributes was nothing more than a duplication of another (that is, both attributes being treated as alike), the effect would be to weight the overall score in favour of the other attribute

Assessments of the creativity of scientific explanations, however, needs more than teacher agreement. **The third question asked about the validity of the assessments.** Amabile (1983a, 1996) took the view that creativity is whatever it means to the assessor but this is not altogether tenable for a teacher who must foster creativity in accordance with generally accepted notions and expectations of bodies such as NACCCE (1999). This means that, for example, the rank order for the Ruler event should be supported to some extent without reference to the assessor's beliefs. There are some clues to the validity of these pre-service teachers' assessments in their responses to particular explanations. It would not be unreasonable for 'correct' explanations to be seen as highly plausible. Given some indication that they were constructed and not simply reproduced by the children, a high score for such explanations was expected. Generally, this was observed. In addition, some of participants saw novelty and plausibility for the child in other explanations, as when the explanation referred to the need to elevate the battery. On the other hand, explanations which were not novel or were irrelevant should have attracted low scores. On average, this was generally the case. The scoring of highly speculative and moderately plausible explanations was expected to fall between these extremes. Again, on average, this was often the case. But, for all kinds of explanation, the variation in scoring was usually large. This points to an underlying, coarse level of validity in the assessments with significant variation between individuals.

What does this mean for classroom practice? These pre-service teachers had completed their training and were about to begin their careers in the classroom where they would be expected to foster creativity across the curriculum. As far as the hypothesis space in science is concerned, their intuitive judgements about the creativity of explanations could be very unreliable indicators of the success or

otherwise of their teaching, of children's creative strengths and weaknesses in science, and of what to do to meet the children's learning needs in the creative context. The responses hinted at some sources of difficulty. A teacher's concern in science can be for obtaining 'the right answer' (Newton & Newton, 2000). Responses to 'correct' explanations revealed that some participants found them irresistible, even when they were reproduced, that is, not novel to the child. Discussion with the groups of participants also revealed that some holistic assessments were largely based on novelty – the more imaginative the explanation, the better it was. After assessing separate attributes of creativity, several of these novices confessed that judging what is plausible in science is difficult. In particular, there was a tension between judging an explanation against a background of adult knowledge and seeing it from the child's point of view, noted in the response to the idea that the battery should be lifted higher.

Nevertheless, there was evidence of something to build on. There was some underlying validity in many responses and agreement improved when attention was focused on specific aspects of scientific creativity. Training could be aimed at helping the teacher become less of an outsider. First, it needs to ensure that the nature of scientific creativity in the hypothesis space is understood, that tentative explanations can be 'wrong' and still be creative (as in the scientist's world) and that creativity involves more than novelty. Talk may not be sufficient; teachers may benefit from tasks which call on them to construct scientific explanations themselves and consider their novelty, plausibility and elegance. Second, they need to see explanations from the child's point of view. Direct experience of and a conscious study of children's scientific explanations in the classroom seems likely to be beneficial. When working

with children, novices may also need to know that fostering scientific creativity means encouraging the deliberate construction of plausible, productive explanations and acknowledging elegance where it exists. They will need to recognise that in the child's world, the 'correct' explanation may not always be the most plausible (given the uncommon sense of science (Wolpert, 1992)). Investigations may lead to the 'correct' explanation or the teacher may have to suggest it as a tentative possibility for testing but the process of getting there need not be at the expense of creative thought.

A few cautionary words may be helpful at this point. Teacher trainers may find a lot of variation in responses from novice teachers. This, of course, can be productive in that it lends itself to discussion. Trainers may also find different responses than those noted here stemming from the different experiences of their trainees. If training is provided for practising teachers, their greater experience with children may also make it easier for them to take a child's point of view. On the other hand, their experience and understanding of scientific creativity in the hypothesis space could be limited (Newton & Newton, 2009b). Novice or experienced, teachers need to be aware that they should also give attention to the processes children use to arrive at an explanation and consider if and how they might be improved. At the same time, teachers will need to recognise that the nature and relative significance of attributes like plausibility can vary with the subject, as in the contrast between art and architecture (Glück, et al., 2002). While there are parallels in creativity in different subjects, there are also differences and these should not be overlooked in integrated approaches to teaching. For instance, while what is appropriate in science can be interpreted as plausibility, it is generally interpreted as function in technology's application space. At the same time, the

weight given to each component may be different in different spaces. Finally, it is also important to avoid the impression that such knowledge and skill will lead with certainty to great precision in assessing scientific creativity. In the context of the arts, Sefton-Green (2000, p. 220) wrote that: 'evaluation, as it is practised, is always much more crude than theorists, even those grounded in practice, want to acknowledge. By definition, what happens in the day-to-day hurly-burly cannot be as subtle as we would like.' This seems an fitting comment to make about the intuitive assessment of creativity in the hypothesis space. Given such caution, however, the assessment may suggest to a teacher that, for example, a particular child rarely produces novel explanations, another's explanations are novel but tend not to be scientific, and another's are novel, scientific yet incomplete or superficial. But this is only of use if the teacher provides opportunities for creative thought and guidance to help these children develop.

Conclusion

Teachers are increasingly urged to foster creativity. Assessing children's creative activity should help a teacher meet children's learning needs. The evidence from this study suggests that an intuitive, holistic approach to assessment in the scientific hypothesis space can have a little validity but is very unreliable. Teachers who use it cannot be sure they have worthwhile evidence to act on. Assessment was found to be a little more reliable when attention switched from a global assessment to a particular assessment of novelty, plausibility and elegance. This particular approach has the further advantage that it supplies information about specific outcomes of creative thought which could point a teacher more precisely to where help is needed. Nevertheless, the variation between these pre-service teachers' assessments of attributes of creativity was often large. Causes may include uncertainty about the nature of scientific creativity in

the hypothesis field, difficulty in taking a child's point of view, and giving credit for non-creative, reproduced, 'correct' answers. Teacher trainers may find it productive to address these in their courses. They will also need to point out that creativity in different domains is not entirely the same, particularly in the relative emphasis given to its attributes. This means that teaching subjects in an integrated way needs great care when the intention is to foster creative thought.

This study focused on the products of creativity or problem solving in the hypothesis space where potential explanations of scientific events are constructed. It remains to be seen if this difficulty in assessment would be observed in the experiment and application spaces and to what extent more extensive experience of children's thinking helps teachers assess scientific creativity with more reliability and validity.

References

- Amabile, T.M. (1983a) *The social psychology of creativity* (New York, Springer-Verlag).
- Amabile, T.M. (1983b) The social psychology of creativity: a componential conceptualization, *Journal of Personality and Social Psychology*, 45(2), 357-376.
- Amabile, T.M. (1996) *Creativity in context* (Boulder, Westview Press).
- Ausubel, D.P. (1978) *Educational Psychology* (New York, Holt, Rinehart & Winston).
- Balchin, T. (2005) A creativity feedback package for teachers and students of Design and Technology in the UK, *International Journal of Design and Technology Education*, 10(2), 31-43.
- Beattie, D.K. (2000) Creativity in art: the feasibility of assessing current conceptions in the school context, *Assessment in Education: Principles, Policy & Practice*, 7(2), 175-192.

- Berg, R. (2009) Evaluating scientific theories, *Philosophy Now*, 74, 14-17.
- Besemer, S.P. & O'Quin, K. (1987) Creative product analysis, in: S.G. Isaksen (Ed.) *Frontiers of Creativity Research* (pp. 341-357) (Buffalo, Bearly).
- Besemer, S.P. & O'Quin, K. (1999) Confirming the three-factor creative product analysis matrix model in an American sample, *Creativity Research Journal*, 12(4), 287-296.
- Boden, M.A. (1996) What is creativity? in: M.A. Boden (Ed.) *Dimensions of Creativity* (pp. 75-117) (Cambridge, Massachusetts, Massachusetts Institute of Technology).
- Boden, M.A. (2004) *The Creative Mind - Myths and Mechanisms* (London, Routledge).
- Claxton, G. (1997) *Hare Brain, Tortoise Mind* (London, Fourth Estate).
- Colman, A.M. (2003) *A Dictionary of Psychology* (Oxford, Oxford University Press).
- Craft, A. (2005) *Creativity in Schools: Tensions and Dilemmas*. (London, Routledge).
- Craft, A. (2008) Nurturing creativity, wisdom, and trusteeship in education, in: A. Craft, H. Gardner & G. Claxton (Eds.) *Creativity, Wisdom and Trusteeship* (pp. 1- 15) (Thousand Oaks, Corwin).
- Cropley, A.J. (2001) *Creativity in education and learning* (London, RoutledgeFalmer).
- Csikszentmihalyi, M. (1996) *Creativity: Flow and the psychology of discovery and invention* (New York, HarperCollins).
- DES (Department for Education and Skills) (2005) *Education and Training Statistics for the United Kingdom* (London, The Stationery Office).
- DES (Department for Education and Skills) (2006) *First Release* (London, DES).
- DfES (Department for Education and Skills) (2003) *Excellence and Enjoyment* (London, DfES).

- Euster, J.R. (1987) Fostering creativity and innovation, *College Research Library*, 48, 287-8.
- de la Harpe, B., & Peterson, J. F. (2008). A model for holistic studio assessment in the creative disciplines, *Proceedings of the 2009 ATN Assessment Conference*, Adelaide, Australia.
- Fox, H. (1963) A critique on creativity in science, in M. Coler (Ed.) *Essays on creativity in the sciences* (pp. 123-152) (New York, New York University Press).
- Gardner, H. (1993) *Creating minds* (New York, Basic Books).
- Glück, J., Ernst, R. & Unger, F. (2002) How creatives define creativity: definitions reflect different types of creativity, *Creativity Research Journal*, 14(1), 55-67.
- Gouge, K. & Yates, C. (2002) Creating a CA programme in the arts: the Wigan LEA arts project, in: M. Shayer & P. Adey (Eds.) *Learning and Intelligence: cognitive acceleration across the curriculum from 5 to 15 years* (pp. 85-102) (Buckingham, Open University).
- Hennessey, B.A. (1994) The consensual assessment technique: an examination of the relationship between ratings of product and process creativity, *Creativity Research Journal*, 7(2), 193-208.
- Hickey, M. (2001) An application of Amabile's consensual assessment technique for rating the creativity of children's musical composition, *Journal of Research in Music Education*, 49(3), pp. 234-244.
- Houtz, J.C. & Krug, D. (1995) Assessment of creativity, *Educational Psychology Review*, 7(3), 269-300.
- Jardine, L. (2000) *Ingenious Pursuits* (Abacus, London).
- Kaufman, J.C. & Sternberg, R.J. (2006) *The International Handbook of Creativity* (New York, Cambridge).

- Keogh, B. & Naylor, S. (1999) Concept Cartoons, teaching and learning in science: an evaluation, *International Journal of Science Education*, 21(4), 431-446.
- Klahr, D. & Dunbar, K. (1988) Dual space search during scientific reasoning, *Cognitive Science*, 12, 1-48.
- NACCCE (National Advisory Committee on Creative and Cultural Education) (1999) *All Our Futures: Creativity, Culture and Education* (London, DfEE).
- Newton, D.P. (2000) *Teaching for Understanding* (London, Routledge).
- Newton, D.P. & Newton, L.D. (2000) Do Teachers Support Causal Understanding through their Discourse when Teaching Primary Science? *British Educational Research Journal*, 26, 599-613.
- Newton, D.P. & Newton, L.D. (2009a) Some students' conceptions of creativity in school science, *Research in Science and Technology Education*, 27(1), 45-60.
- Newton, D.P. & Newton, L.D. (2009b) What teachers see as creative incidents in elementary science lessons, *International Journal of Science Education*, iFirst, 1-17.
- Petty, G. (1997) *How to be Better at Creativity* (London, Kogan Page).
- Richards, R. (1993) Everyday creativity, eminent creativity and psychopathology, *Psychological Inquiry*, 4(3), 212-217.
- Roberts, S. (2004) Self-experimentation as a source of new ideas, *Behavioral and Brain Sciences*, 27, 227-288.
- Rogers, M. & Fasciato, M. (2005) Can creativity be assessed? Paper presented at the British Educational Research Association Conference, University of Glamorgan, 14-17 September, 2005.
- Rose, J. (2009) *The independent review of the primary curriculum: final report* (Nottingham, DCSF Publications).

- Rust, C. (2002) The impact of assessment on student learning, *Active Learning in Higher Education*, 3, 145-157.
- Sefton-Green, J. (2000) From creativity to cultural production: shared perspectives, in J. Sefton-Green and R. Sinker, *Evaluating Creativity* (pp. 216-232) (London, Routledge).
- Siegel, S. (1956) *Nonparametric Statistics* (Tokyo, McGraw-Hill Kogakusha).
- Tatarkiewicz, W. (1980) *A history of six ideas: an essay in aesthetics* (The Hague, Martinus Nijhoff).
- Treffinger, D. & Poggio, J. (1972) Needed research on the measurement of creativity. *Journal of Creative Behavior*, 8, 101-106.
- Treffinger, D.J., Young, G.C., Selby, E.C. & Shepardson, C. (2002) *Assessing creativity: a guide for educators* (Storrs, The National Research Centre on the Gifted and Talented).
- Williams, F.E. (1970) *Classroom Ideas for Encouraging Thinking and Feeling* (Buffalo, DOK Publishers).
- Wolpert, L. (1992) *The Unnatural Nature of Science* (London, Faber).

Appendix

Assessing Creative Scientific Thinking: Sound

A teacher has Year 1 class children (6 years old) cup their hands behind their ears so they note an increase in loudness of sounds. She asked them to explain why it made the sounds louder. Here are some of the explanations:

- a. I'm catching the noises with my hands and making them go in my ears.

- b.** I've got bigger ears now.
- c.** It stops noises from behind getting in.
- d.** It opens up your ears.
- e.** Our hands catch the noises that matter.

Assessing Creative Scientific Thinking: Ecological niches

A teacher had some Year 3 children (8 years old) look at tree trunks. They noted that the shady side of each trunk had green stuff growing on it. The teacher asked them to explain why the green stuff grew only that that side of the trunks. Here are some of the explanations:

- a.** The green stuff hides itself away so nothing can eat it.
- b.** It must like it there.
- c.** It must have been rubbed off the other side by people leaning on the tree.
- d.** The green stuff must not like to get hot so it stays on the shady side.
- e.** It grows up out of the ground on that side.

Assessing Creative Scientific Thinking: Electricity

A teacher asked some Year 3 children (8 years old) to help her make a torch bulb light up. She followed their suggestions. They had her connect a wire from the top of a battery to the bulb. The bulb stayed off. She asked them to explain why. Here are some of their explanations:

- a. The battery's too old. I've seen that happen before with my torch.
- b. The electric's got stuck in the wire.
- c. The electric has to go in *and* out of the bulb. That lets it only go in.
- d. The battery is warming up – it takes time.
- e. You need to hold the battery higher so the electric can run down to the bulb.

Assessing Creative Scientific Thinking: Dissolving

A teacher shows some Year 4 children (9 years old) stir a sugar cube in water until it dissolves entirely. S/he asked the children to explain what happened to the sugar cube.

These are some of the explanations:

- a. It whirled around until none was left.
- b. Bits came off and broke up smaller and smaller until we just couldn't see them.
- c. The water soaked into the sugar cube and now they're the same.
- d. Too much sugar clogs things up.
- e. There are secret holes in the water and we stirred the sugar cube into them

Assessing Creative Scientific Thinking: Forces

A teacher has some Year 6 children (10 years old) watch her hold a wooden metre rule by one end so that it hung vertically with the other end about 50 cm above a hard, concrete floor. She dropped the ruler, the lower end hit the floor and the ruler bounced up and she caught it. She asked the children to explain why the ruler bounced. These are some of the explanations:

- a.** First, there's gravity; then there's antigravity. Gravity makes it go down then antigravity takes over and makes it come up.
- b.** The Earth is like a big rubber ball and the ruler bounces off it.
- c.** The ruler is made of wood. I bet we could make a wooden ball that bounced.
- d.** I've seen it in a book in the library. It said the ruler buckles a bit then it jumps back to normal in the air.
- e.** If you look at it, the ruler bends a bit when it hits the floor so it looks like a bow. I think that when it straightens, it pushes itself up off the floor.

The second form of the assessment included the explanations, an explanation of the terms, instructions, scoring rules and the following grid (Figure 3).

<<Figure 3>>