

Assessing Students' Creativity: Lessons from research

Dr Tom Balchin, Brunel University

WHY ASSESS CREATIVITY?

There are many valid and important reasons for being concerned with creativity assessment. For example, I have identified ten general roles for creativity assessment which include:

- Helping to remove creativity from the realm of mystery and superstition;
- Helping students to recognise their own strengths and talents; enabling people to know and understand themselves better;
- Expanding students' risk-taking parameters;
- Helping teachers to discover unrecognized or untapped potential/talent;
- Providing baseline data for assessing individuals or groups and results which can guide teachers in planning and conducting appropriate and challenging instruction;
- Highlighting current educational problems such as 'marking to the assessment criteria'. Consistent gap between a high score in hitting assessment criteria yet low creativity will tell their own stories.
- Helping in the recognition and 'reward' processes for students;
- Providing a common language for communication among professionals about the nature of creative abilities and skills;
- Giving a boost to students who tend to fall outside society's norm behavioral standards and are judged badly as a result of poor academic work;
- Providing pre-test and post-test data for group comparisons for further researcher evaluation;

HOW?

Creativity assessment efforts might be qualitative, quantitative, or both. Analysing qualitative data is a process which considers relevant contextual issues, possible biases, and values; it is concerned more with discerning the meaning of information rather than with formulating and testing statistical hypotheses, although there exists possibilities of statistical scores for creativity through mathematical means. An example of qualitative data analysis however, is an observer's description and analysis of a student's curiosity and creativity, as expressed in spontaneous exploratory behaviour in a typical classroom room setting. Data might be gathered in classrooms, in the workshop, and even on the playground, involving many instances and examples of the student's curiosity and exploration, gathered over a period of several weeks.

The complex and multidimensional nature of creativity cannot be captured effectively and comprehensively by any single instrument or analytical procedure. Systematic efforts to understand creativity require a well-planned process of studying individuals or groups, including both qualitative and quantitative data. **Creativity assessment might be regarded as an attempt to recognize or identify creative characteristics or abilities among people, or to understand their creative strengths and potentials.** Measurement might play a specific role in creativity assessment to the extent that specific tests, inventories, or rating scales provide evidence to help answer such questions. We would be concerned with creativity assessment in education, for example, if we were to pose such questions as:

- Who are the most (or least) creative students in this class?
- What characteristics suggest that a particular student is very creative?
- What are the creative strengths of the people in this group?
- What is the best climate for creative outcomes?

- How is creativity expressed differently among individuals of varying learning styles or preferences?
- How is creativity expressed through products?
- How might we optimize a group's performance, or design the most effective training experience for a team or work group?

Measurement commonly plays an important role in evaluating instruction or training efforts related to creativity. If a special problem for students purported to enhance or stimulate students' creative thinking skills, for example, pre- and post-tests might be used as part of an evaluation design. The kinds of questions posed might include, for example:

- Was the programme effective in enhancing students' creative thinking and problem solving skills?
- What impact did the programme have on those who participated in it?
- Were participants better able to recognize problems, generate ideas, and plan for creative action after the training than they were prior to it?
- Did participants in an experimental group demonstrate greater gains in creativity than students in a control group?

WHAT?

Although there has been growing emphasis on the importance of creativity and thinking skills in education, it does not seem that systematic efforts to measure creativity and problem solving are common at all, whether in relation to assessment of students' characteristics and needs or to evaluation of instructional or programmatic outcomes. Teachers might at best compare previous products or motivation levels. Consensual agreement amongst teaching staff concerning individuals is possibly going on, but there is no single instrument which might be used to assess all the relevant and important aspects of creativity. Validity is out of the window! However, many instruments have been created to assess various aspects of creativity; one recent review for example, identified more than 200 tests, inventories, rating scales, and checklists that purported to measure some aspect of creativity or its correlates (Isaksen et al. 1993). These instruments vary considerably in their appropriateness or usefulness for subjects of varied ages or across different settings. In many cases, the evidence for their ability and reliability is incomplete or not fully satisfactory.

At best, it would be necessary to develop a complex composite of several instruments for any particular assessment or evaluation context, based especially around the **4 P's** (product, process, person, and press (promoters/providers)) and even then, results must be qualified and generalisations made with great caution. Measuring creativity is not an easy task; measures have often been administered inappropriately, used to assess skills or abilities they were not intended to measure, or interpreted inappropriately.

For Processes

There exists another useful way to measure creativity; to simply 'ask the subject'. This is not a profound position, but yet the procedure is rarely used. The predominant preference in the field today is to identify creativity by indirect methods (i.e., predictors) that essentially have little to do with the real criteria of creativity. Asking the subject has an important advantage. The critical incidences that occurred during the creative process will be known better to the individual than to any observer. Indeed, earlier it was argued that other individuals can not always discriminate creativity from their own general opinion of the subject. On the other hand, the subject himself should have a good idea of his creative ability in a wide variety of areas, and especially the moment of inspiration that caused him/her to take creative action. I have found that when compared to observer ratings and other assessment procedures, self-reports have been found to be superior in the measurement of many psychological traits. The CFP includes creative moments sheets, and these have been very successful in getting students to reveal creative thoughts.

For Products

Here's the technique that my studies have shown is has most internal validity. The technique caused the same answers (within spaces of 3) to be given to criteria of creative products with 6 groups of teachers (to the order of .83 on the Pearson Correlation scale. An experiment that achieved this is as follows:

15 Heads of Design and Technology from secondary schools in Greater London were asked to bring in one product that a student had made at GCSE in resistant materials, textiles, systems and control or graphic products. Food technology (part of the current curriculum) was deemed inappropriate for the experiment! They were given product criteria sheets, worked out as part of a creativity feedback package (Balchin 2005).

7 products were brought by the teachers:

- *full-size adjustable garden chair*
- *3D cinema promotional poster*
- *small jewellery box/cabinet*
- *aluminium lamp*
- *cultural choker (in a variety of metals)*
- *plastic key-ring*
- *a dress (half-finished).*

I also supplied 3 products, made by students in schools taking part in the Goldsmiths College DFE/QCA Assessing Innovation project:

- *a hat with blue water in the top to mimic a cooling pack (Textiles)*
- *a drawer/box with black suede (RM)*
- *a small barometrical device (S&C).*

The teachers were all given 30 minutes to assess them individually, then consensually in small groups according to the criteria on the CFP product page. The CFP is a creativity feedback package designed for teachers and students of design and technology designed round the 4 P's. The product part has 7 criteria to be scored on a 12-point scale:

- *uniqueness*
- *associations of ideas*
- *risk-taking*
- *potential*
- *operability*
- *well-craftedness*
- *attractiveness.*

Four criteria described the creative *concept*, or idea, and 3 criteria describe the *quality of build*, which evaluate how well how the creative thoughts have been brought into the 'made world'. The emphasis in the CFP's product sheet is that creativity is seen in *both* the concept and the standard of build that the resulting product shows. But it is the *concept* stage where the unique ideas are brought forth, and the *product* stage is the manifestation of those creative ideas. The latter cannot occur without the former. The 'quality of build' is a *vehicle* for the creativity.

These criteria were subsumed from an exhaustive list of qualities of creative products by research and trialling in pilot schools. However, *numbers were not critical to this exercise*, because the scorer will have his/her unique scale of judgement and frames of reference to the product. The scores *therefore have no real meaning as numbers*; they only help the scorer to get to grips with the criteria; and reflect the way in which individual thinking processes can be changed by magnifying awareness of creative processes.

The forms were filled in by each group in ten minutes of discussion whilst passing the product between them, and once analysed, each of the 6 groups agreed consensually about the creativity of products to the order of 83.3% (if agreement constitutes scores within 3 places of each other within a 12 point scoring system). Furthermore these were done without reference to any accompanying portfolio, knowledge of the process, personal information about the maker or knowledge of the particular climate/environment it was constructed in. There is an adjunct to this: as the product gets more unique, the groups' abilities to agree on product creativeness seems to lessen.

A Consensual Definition of Creativity

After an exhaustive literature review and an empirical study of the assessment of creativity, I am convinced that consensual expert assessment is probably the best way to achieve a high degree of validity in evaluation. **This concerns the idea that a product or response is creative to the extent that appropriate observers independently agree it is creative.** Appropriate observers are those familiar with the domain in which the product was created or the response articulated. Thus, creativity can be regarded as the quality of products or responses judged to be creative by appropriate observers, and it can also be regarded as the process by which something so judged is produced. Current research has shown that any identification of a thought process as creative must finally depend on the fruit of that process—a product or response. Similarly, even a clear specification of the personality traits that mark outstandingly creative individuals would have to be validated against their work. A product-centered operational definition is therefore clearly most useful for empirical research in creativity.

CAN A PRODUCT BE A CHANGE?

Perhaps the most important feature of this definition is its reliance on subjective criteria. There must be particular characteristics or attitude statements of persons or products that observers systematically look to in rating them on scales of favourability or physical attractiveness or creativity, but, ultimately, the choice of those characteristics seems to be personal to the evaluator. As this is the case, any use of such scoring must mean that this assertion of the necessarily subjective nature of creativity assessment must be underscored.

The consensual definition conceptually identifies creativity with the specific products under investigation. It may indeed be possible to identify particular objective features of products that correlate with subjective judgments of creativity or to analyse the nature of subjective correlates of those judgments, but this definition makes it unnecessary to attempt to specify those objective features or the characteristics of those subjective reactions beforehand.

RELIABILITY AND VALIDITY OF CONSENSUAL ASSESSMENT

The most important criterion for the results of this assessment procedure is that the ratings be as reliable and valid as possible. Reliability is about repeatability, and validity is about truth. By definition, interjudge reliability in this method is equivalent to construct validity; if appropriate judges independently agree that a given product is highly creative, then it can and must be accepted as such.

In addition, it should be possible to separate subjective judgments of creativity from judgments of technical goodness and from judgments of aesthetic appeal. It is fairly clear that for some domains of endeavour it may be relatively difficult to obtain ratings of aesthetic appeal and technical quality that are not highly correlated with ratings of creativity. However, it is important to demonstrate that it is at least possible to separate these dimensions. The discriminant validity of the measure would therefore be in doubt; judges might be rating something as creative simply because they like it or because they find it to be technically well done.

Judges' ratings can be used to determine if the original task presented to subjects was appropriate for the purposes of a social-psychological methodology. Certainly, if virtually all of the subjects in a random sample of a population are able to do the task and report no technical difficulty in doing so (i.e., in manipulating the materials, in finishing within a reasonable period of time, and so on), this suggests that the task was well-chosen for these purposes. If later judging of the products reveals a low correlation between judged creativity and experience-related characteristics of the subjects (e.g., age, experience with the particular type of materials), then the task can truly be considered a satisfactory one.

However it seems that appropriate judges must be chosen; as I have indicated, judges should be familiar with the domain of endeavour in which the product was made. Furthermore, it is important that creativity in the evaluation of products tests correlate with some measurable quality. In this way consistency can be kept and certain characteristics correlating to 'creativity' can continue to be identified.

References

Balchin, T. (2005) "A Creativity Feedback Package for Teachers and Students of Design and Technology in the UK." in the *International Journal of Design and Technology Education*, **10**, 2, p. 31-43.

Isaksen, S.G., Murdock, M.C., Firestien, R.L. and Treffinger, D.J. (Eds.) (1993) *Understanding and Recognising Creativity: the Emergence of a Discipline*. Norwood, N.J.: Ablex, p. 34-99.