Submitted to AARE November 1994

AN EVALUATION OF THE EFFECTS OF THE
CRITERION - BASED ASSESSMENT PROCESS ON DESIGN OUTCOMES
AND STUDENT PERFORMANCE.

Judy Hodgman,
Department of Secondary and Post Compulsory Education,
University of Tasmania.

Abstract

Design assessment criteria which reflect the educational objectives (intended learning outcomes) of units and programs are important tools for teaching the subject. The criteria are used to assess students' work and are made available to years 11 and 12 students at the beginning of each design project. These results, in turn, reflect the educational objectives of the course.

The aim of the research discussed here was to test the hypothesis that assessment criteria have the capacity to direct the design decision making and problem solving process, and directly influence design outcomes. The research investigated the extent to which the student used assessment criteria and the effect that this had on the development of normative and summative evaluation skills.

The outcomes of students' projects that had, and those that had not been given assessment criteria, were compared. The results showed that students who had no access to assessment criteria achieved a statistically significantly better result overall for design and problem solving and they demonstrated a greater use of formative and summative evaluation skills. The research concluded that (a) the criterion based assessment process, which provides evaluative criteria at the outset, limits the development of essential design evaluation skills and (b) that the national evaluation process has doubtful validity in terms of linking course objectives to learning outcomes.

Introduction

Teacher and student evaluations of learning outcomes is an important part of teacher education. Assessment criteria and course objectives are linked and provide the rationale or the course in terms of process and content.
A minimum requirement of the design and technology teacher is that they should be capable of evaluating their own work as well as that of others. It was intended that the introduction of the criterion based assessment process would overcome concerns expressed by Scriven (1991) that few teachers have any idea how they or others can or should evaluate their own work or that of others.

Teaching or evaluation that focuses on identifying the competencies needed by the trainee to various subjectively determined achievement levels is now in question (Scriven 1991, p.84).

Evaluation training, according to Scriven (1991) is the training of the citizenry in evaluation technique, and is the only satisfactory long-run approach to improving the quality of our lives without extraordinary wastage of resources.

The new Australian Design and Technology curriculum is concerned with developing skills in design and problem solving. The amount of problem solving in design project work is often dependant on the degree of qualification used by the teacher to modify the original problem or in the number of limitations or restrictions used to direct course content. Problem solving requires the use and understanding of qualitative evaluations which are required equally by student and teacher. It is not enough to teach students how to represent a problem and how to do it, they need also to learn how to monitor and evaluate what they are doing.

This process is not a simple one. In the usual taxonomy of cognitive processes it is listed as the most sophisticated of all, and its logic is complex enough to have evaded satisfactory analysis for two millenia (Scriven, 1991, p.4).

A substantial part of good evaluation of products is both formative and summative as well as 'reactive effective'. As Scriven (1991, p 169 ) put it:`
Formative evaluations should take the form of an analytical evaluation avoiding any other obligations that could bog it down or expectations that could result in disappointment.

'Reactive effective' has two components: content effects and process effects. The success of this type of evaluation depends on both. The internal assessment criteria used to assess students' performance during the process of designing, making and appraising is concerned with process and performance. The external assessment criteria assesses design outcomes (products) and is concerned with the quality of content. Quality assurance and quality control originated in the manufacturing and engineering areas. This kind of evaluation is usually internal and formative, and is the kind of formative evaluation that is essentially early-warning and summative, ensuring that the product, when it reaches the consumer, will be acceptable.

Formative evaluation is contrasted with summative evaluation and is nearly always internal to the person making the evaluation. It is conducted during the development or improvement of a product or program and often more than once with the intent of improving the outcome. It should begin with preformative evaluations conducted during the design phase. The difficulties that often present themselves in the design outcome are often due to poor preformative evaluations. This initial evaluation process determines the design criteria needed for the successful implementation of the design process.

The research, therefore, needed to gain an understanding of how students used assessment criteria during design and problem solving phases and the effect that these criteria had on product outcomes and student performance. The importance of designing a relevant test was recognised. It had to include student involvement in both 'quality assurance' and 'reactive effective' evaluations. Tyler's 1934 generalized technique for constructing achievement tests was used and is manifest in the framework that follows below.

The Research Process

1) Identifying objectives of the educational program. This was accomplished by listing the objectives or learning outcomes of the design project work.
The project work was assessed using the eight most commonly used internal assessment criteria and the two most commonly used external assessment criteria at National Certificate level (TCE); which are the two post-compulsory years following eight years of compulsory schooling.

2) Defining the objectives in terms of behaviour and content. Specific skills development and course content knowledge were listed under constraints (limitations) in terms of materials and processes.

3) Identifying situations where the objectives are utilised. Identifying the human need or a human situation. A statement of the human need was outlined in the design brief which asked the student to respond to this specific human need.

4) Defining or identifying the process. An appropriate problem solving design process was used. It typically included - identification and investigation of the problem, creating testing and evaluating design solutions to the problem.

5) A record of events. A log was kept which listed the comments and observations made during the four design phases. Students' progress and performance throughout these phases was assessed using the predescribed assessment criteria (group A).

6) Internal and external appraisals. These appraisals were done using a nationally agreed assessment process. This included an internal and external assessment of the project work. The external assessment involved the collaboration of two internationally acclaimed product designers who were not involved with project work and who had no contact with the students throughout the process phases.

7) A representative sample of students work. A selection was achieved using quasi-experimental design methods. This method is described under 'procedure' in the research section.

Campbell and Stanley (1963) argued that quasi-experimental design can prevent subject reactivity
from influencing outcomes and that serious evaluations should try for it whenever true experiments are impossible. Quasi-experimental design is the next best method to use when a random allocation of subjects for the control and experimental groups cannot be arranged (Scriven, 1991, p.297).

The research used an interpretivist/constructivist approach to collect and analyse data. Discussion and descriptions provided data for responsive evaluations and this was used in conjunction with the quasi-experimental design test results described below. This combined approach provided key links between the use of the criterion based assessment process, student performance and design outcomes.

Human behaviour is important to this type of research as the phenomena of reactivity cannot be disregarded. Humans are not inert and are capable of many interpretive actions. The research involved 'responsive evaluation' which picks up on whatever occurs and deals with it as it seems appropriate.

The emphasis is on personal experiences and rich description and recognises that evaluation data derived from constructivist enquiry have neither special status nor legitimation and simply represent another construction to be taken into account in the move toward producing knowledge claims (Guba and Lincoln, 1989, p.43).

Procedure

A four-phase quasi-experimental research program was devised using a control group A and an experimental group B. Thirteen (13) students made up the control group A and seven (7) students the experimental group B. The B group was paired with seven students in the A group.

A total of 20 students participated in the two design projects over two separate 7 week sessions and all were presented with the same design problem. The stratification of the two groups reflected socio-economic status, experience, age and gender.

Both groups were introduced to the design problem through a 'design brief' which outlined the human need. It asked students to respond to this need by finding one or more solutions to
the problem and articulating their ideas in two and three dimensions. Group A were also given additional information in the form of the 10 national assessment criteria used in Technology education.

The research focussed on differences that emerged in both content and process during the designing of a product in response to the design brief and began by synthesizing the assumptions of the students.

Phase 1  Mapping Design Assumptions

This phase collected ranked preformative design evaluative criteria from both groups. Interviews were conducted throughout the phase and comments recorded. The ranking of preformative evaluative criteria allowed participants to respond to a specific design problem necessary for the design process to occur in phase 3.

Phase 2 - Synthesising Design Assumptions

Preformative design evaluation criteria from both groups developed in phase 1 were ranked and design ideas were articulated in two dimensions (design development drawings). These were further developed for use in phase 3.

Phase 3 - Experimental Pedagogical Applications

All students were asked to use the design development drawings from phase 2 for the transition from 2 dimensions to 3 dimensions. They were asked to present their final solution to the design problem in three dimensions.

Phase 4 - Teacher And Expert Evaluations.

Students' performances were analysed and assessed by the researcher using the 8 pre-described internal assessment criteria. Two experts in the field of product and industrial design were invited to assess the products presented to them based on the problem solving skills evidenced in the final solution (2 external assessment criteria). These assessments were made using quantitative and qualitative product evaluations and consensus was reached using an
ordinal scale (ranking and or grading) where each product was compared with the rest.

Results

The interim results of observations and interviews conducted during the two design sessions follow.

Phase 1 and 2

Group A (10 assessment criteria made available at the start of the project)

1) A large proportion of student time was spent questioning and translating assessment criteria, prior to any preformative design evaluations being made. Teacher's interpretations of the assessment criteria were constantly sought.

2) All 13 students commented on the need for this information which they thought necessary to help them direct their design process.

3) They all agreed that the teacher's expectations and assessments of them and their work was more important than their own evaluations.

4) 11 students thought that the design evaluations that they had made about the original design problem were important to the final product but not to their final results.

5) Most students encountered problems with decision making when they had to rank the preformative design evaluative criteria. They requested summative evaluations from the teacher throughout the duration of the project.

Phase 1 and 2

Groups B (experimental group)

1) No time was spent in translating assessment criteria or defining teacher expectations.
2) All students agreed that they would have liked the assessment criteria and saw them as advantageous to improving their mark but not to improving their design solution.

3) All students agreed that their ranked design evaluative criteria were the most important part of the problem solving process and that this was necessary in order to guide the project work through to phase 3.

4) More time was spent on this evaluative phase by the B group than by the A group.

5) A number of students recognised the importance of evaluation skills in determining the design solution (product).

Phase 3

Groups A

Two and three dimensional design work was undertaken, with all students presenting a final solution to the problem within the time limit.

1) Most students agreed that they should have been more conscientious about using the ranked design evaluative criteria and some students admitted to losing them.

2) All students agreed that they needed to accommodate the assessment criteria which reflected the teacher's expectations which they considered to be important to their final result.

3) All students articulated their ideas in three dimensions but only to mock-up stage.

4) Most students accommodated the assessment criteria demonstrating their understanding of the criteria. They all agreed that their goals were to satisfy these assessment criteria.
5) The ability to perform practical tasks varied slightly within the group.

Phase 3

Group B

Two and three dimensional design work was undertaken with all students presenting a final solution to the problem within the time limit.

1) All seven students agreed that they used their own ranked design evaluative criteria to develop their own ideas about the project.

2) All students agreed that they continuously evaluated their own work and performance as well as other students.

3) All students produced at least three mock-ups and one final working model with 4 students attempting working prototypes.

4) All students agreed that they needed to set their own goals.

5) The ability to perform practical tasks varied slightly within the group.

Results of the internal and external assessments

Students with access to assessment criteria totalled 13 in group A and students with no access to selected criteria totalled 7 in the experimental group B and are identified by (*) in table 1

Table 1 Internal and External Assessments

<table>
<thead>
<tr>
<th>Student Number</th>
<th>Internal</th>
<th>Expert</th>
</tr>
</thead>
<tbody>
<tr>
<td>1)</td>
<td>NANA</td>
<td></td>
</tr>
<tr>
<td>2)</td>
<td>HANA</td>
<td></td>
</tr>
<tr>
<td>3)</td>
<td>OASA</td>
<td></td>
</tr>
<tr>
<td>4)</td>
<td>SANA</td>
<td></td>
</tr>
<tr>
<td>5)</td>
<td>SANA</td>
<td></td>
</tr>
<tr>
<td>6)</td>
<td>OASA</td>
<td></td>
</tr>
<tr>
<td>7)</td>
<td>HANA</td>
<td></td>
</tr>
</tbody>
</table>
9) OASA
10) HAOA *
11) OHAHA
12) SAOA *
13) HANA
14) SAOA *
15) SAOA *
16) OA NA
17) HAOA *
18) HAOA *
19) SANA
20) OAOA

Scoring
0 equal to NA or No Satisfactory Achievement.
1 equal to SA or Satisfactory Achievement.
2 equal to HA or High Achievement.
3 equal to OA or Outstanding Achievement.

There were significant differences between the grades awarded for each of the two student groups and between the two assessment mechanisms.

The internal assessments for group A (a mean rating of 2.231) proved significantly better than the internal assessments for group B (a mean rating of 1.571).

The opposite occurred for the external assessment with group A proving significantly worse (a mean rating of 0.651) than group B (a mean rating of 2.875, Table 2). The results of this scoring is graphically portrayed in figure 2.

Effective, reactive, formative and summative evaluations were used by the two groups during the four phases of designing and making, and has been compiled into the following Table.

Table 3

Discussion

The reason for the variation in the two sets of results can be attributed to the evaluation methods used by the two groups. The experimental Group B was primarily concerned with using formative evaluations and this was reflected in their higher score. The reasons
for this may be due to the students need to constantly evaluate their own work and personal performance in the absence of any assessment criteria or teacher interpretations of the assessment criteria. They demonstrated the use of 'reactive effective' evaluations using both formative and summative evaluations. Content effective evaluations were very much in evidence in the finished design, and process effective evaluations were used to help direct the design process.

In contrast, Group A were unsure of their role as evaluators and produced fewer preformative design evaluative criteria. They were preoccupied with teacher expectations of the assessment criteria. They did not demonstrate the use of 'reactive effective' evaluation, the main problem for them being the assessment criteria and not the design problem. Only partial use of 'quality control' was evidenced in their finished design.

These poor preformative evaluation performances may have been due to including too many external considerations such as assessment criteria and specific teacher expectations which, as Scriven (1991) suggests, should be avoided as they can limit or increase expectations which may result in disappointment. The consequence of disregarding the importance of 'content effective' evaluation by this group further encouraged evaluation limitations and may explain their low external mark. In comparison the experimental group B monitored and evaluated their own work throughout the project using 'content effect' and 'process effect' evaluations. This quality assurance or quality control may have resulted in their success in terms of answering the design problem and gaining a higher external mark. According to Silverman (1985), a balance of power is the main separating factor between quality of outcomes and this balance or power requires active participation by those doing the work and those assessing the work. An imbalance of power may have prevented Group A from taking control over much of the decision making required during the project and may have contributed to the lack of quality control evidenced in the final product and in their final results.

Conclusions and Implications

Disparities in outcomes related to the two methods of assessment exist. Where internal assessments alone are used there is strong evidence to support the use of predescribed assessment criteria.
Where both internal and external marks are needed there is strong evidence to support the withholding of assessment criteria at the outset.

The results also confirmed that the internal assessments of both groups could not be used to predict an external result. Little relationship between the two assessment mechanisms was found. Comparison between the two groups would have to be considered separately for each assessment mechanism.

The reasons for this disparity is evidenced in the types of evaluations made by the students during the four phases of the project and are outlined in Table 2.

Concluding Note

The findings presented here cannot claim that the application and primary use of the national assessment criteria is interfering with the development of essential evaluation skills and hindering the learning process. What it does claim is that (a) there is enough evidence to suggest that the criterion based assessment process has suspect validity in terms of defining the main objectives of technology education and (b) that the differences between national evaluation and expert evaluation are unrelated. This raises significant questions as to why these processes have been adopted for determining awards in design and technology education when performance, as well as product outcomes, may be limited by their use. The findings suggest that assessment criteria should be withheld from students to encourage the development of design evaluative skills.

Acknowledgements

This research was funded by the School of Education, University of Tasmania, Launceston.

References


California.


