

1997

## The importance of assessment procedures to student learning outcomes in religious education

Phillip Cox  
*Edith Cowan University*

John Godfrey  
*Edith Cowan University*

---

### Recommended Citation

Cox, P., & Godfrey, J. (1997). The importance of assessment procedures to student learning outcomes in religious education. *Australian Journal of Teacher Education*, 22(2).  
<http://dx.doi.org/10.14221/ajte.1997v22n2.5>

This Journal Article is posted at Research Online.  
<http://ro.ecu.edu.au/ajte/vol22/iss2/5>

## THE IMPORTANCE OF ASSESSMENT PROCEDURES TO STUDENT LEARNING OUTCOMES IN RELIGIOUS EDUCATION

Philip Cox and John Godfrey

Edith Cowan University

Churchlands

### BACKGROUND

The teaching of religious education has failed, in the past, to utilise the basic principles of good education. The use of assessment and evaluation techniques have been shunned by many teachers. The rejection of such essential teaching strategies has tended to be based on theological ideals rather than sound educational arguments.

During the 1980's researchers in the field of religious education were able to present strong arguments that supported the use of assessment strategies in the teaching of religious education. A sound philosophical framework to support the introduction of assessment strategies in the teaching of religious education was developed. This paper tests the validity of the philosophical arguments supporting the use of assessment strategies in the teaching of religious education in a classroom situation. This paper attempts to answer those who continue to challenge the validity of the of the arguments supporting the use of assessment strategies.

Many teachers continue to reject the use of assessment strategies in the teaching of religious education on the basis that religious education is somehow different to the teaching of other subjects.

In Perth, Western Australia, the teaching of religious education has not in the past utilised summative assessment procedures as a tool. Students therefore had no experience of testing in religious education. This lack of exposure to testing in religious education classes created an ideal situation in which to set up an experiment to ascertain the importance of assessment procedures to student learning.

### Assessment and Evaluation: Aspects of Teaching

The process of utilising assessment and evaluation within the context of education relates to the principles of good teaching and classroom management. Assessment and evaluation are necessary aspects of the teaching process because educational objectives are often very broad in their scope and, as such, are often vague. The teacher must therefore interpret these broad objectives and establish specific and tangible objectives. This step enables the teacher to discover if aspects of the subject have been taught. Specific, tangible objectives can be measured, observed, tested, assessed.

The importance of measurement, assessment and evaluation techniques to the teaching process relates to the reason for the process of teaching itself. One assumes that students will be different after a unit of work has been

taught. The question arises as to the degree of difference. Hence measurement, assessment and evaluation are important to determine the degree of difference. Within this context, the main purpose of classroom instruction is to enable students to achieve intended learning outcomes. In so doing the teacher becomes a predictor. The teacher needs to decide to utilise a particular technique because the teacher expects that the chosen technique will be more effective in producing the desired outcome. Selecting a particular technique instead of another requires evaluation of the technique chosen and thus the need for assessment arises. The teaching process requires that assessment and evaluation occur. In this way assessment is not a post teaching procedure, but an integral part of the teaching process.

Cole and Chan (1987) are particularly wary of teachers who are overtly confident of their capacities to make informal judgements about a student's abilities and achievements. They classify this type of teacher as a 'self-reliant assessor' (p. 295). They point out that teachers who shy away from assessment and evaluation strategies on some philosophical ground or principle believe that they can answer the questions relating to effective teaching without utilising the vast wealth of objective information that can be gained through the use of effective diagnostic, formative and summative evaluation.

Religious education teachers in Catholic schools are generally teachers in another

subject area. In their primary teaching area these teachers would never consider abandoning the use of assessment and evaluation strategies, and yet, while holding this firm conviction, easily ignore sound educational strategies when they walk into the religious education classroom.

#### **AIMS OF THE STUDY**

The aim of this study is to investigate whether the use of formal assessment procedures in the teaching of religious education has an effect on student learning outcomes. The Catholic Education Office of Perth has accepted the arguments that the researchers during the 1980's put forward and directed that religious education teachers utilise formal evaluation and assessment strategies. Many classroom teachers were not convinced and continue to challenge the validity of the arguments put forward supporting the use of assessment strategies in the religious education classroom.

#### **SUBJECTS**

Initially eight religious education teachers were involved in the study. Four classes were randomly selected to represent the experimental group. One of the four control class teachers withdrew support for the study part-way through the experiment,

leaving only three classes to represent the control group. Given that 77 students, from three separate classes remained in the study the loss of one class was not seen as detrimental to the outcomes of the study. The experimental group contained four classes totalling 83 students.

The subjects were 160 students in Year 8 (the student's eighth year of formal education) in a metropolitan Catholic high school in Perth. Apart from ensuring a gender balance the students had been randomly allocated to each class.

**DESIGN**

A nested experimental design was utilised to provide the necessary data and to draw conclusions to answer the research questions. There are two levels of effect within the study. The individual class variation in scores is nested within the variation of scores between

the experimental and the control groups. Factor A represents the treatment or non-treatment of the respective groups of students and is the first level of analysis. At this level the two groups include the Experimental Group and the Control Group. The Experimental Group experienced a range of formal assessment procedures (treatment). The Control Group did not experience this treatment. Factor B, at level 2, separates the experimental and control groups into their individual classes. At this level variation of test scores between individual classes is the focus of the analysis. The experimental design is shown in Figure 1.

Figure 1: Nested design of the study

Level 1	Experimental Group	Control Group
Factor A	(Treatment – Formal assessment procedures)	(Non-Treatment - No formal assessment)
Level 2	Class 1-4	Class 5-7
Factor B	(Teacher differences)	(Teacher differences)

**KNOWLEDGE TESTS**

To ensure consistency of scoring of the knowledge test it was decided that a twenty-item four-choice multiple choice test would be used. Through a series of pilot studies in other schools the test items were gradually refined to produce effective distractors. While in some items more than 25% of the students scored the correct response the average item difficulty for this group remained very near 25%. This is well within the range of 20% to 80% set by Kubiszyn and Borich (1987, p. 29).

**RELIABILITY AND VALIDITY OF THE KNOWLEDGE TEST**

The knowledge test was found to be reliable and valid. Internal consistency was tested using a split half reliability index. An odd-even, split-half reliability index of .82 was obtained for the knowledge test.

A discrimination index for each of the twenty knowledge question was also calculated to indicate the reliability of individual items. To determine this index, the upper and lower group boundaries were set at 27%. The average discrimination index is .34.

The post-test and follow-up test design of this study enabled a calculation of a stability reliability index. Given that there was no intervention between these two tests, the reliability index was calculated using these two tests. The time span between the posttest and the follow-up test was two weeks. This analysis produced a Pearson  $r$  of .87, indicating a high degree of similarity between the scores on each test occasion. This result indicates that the knowledge test is reliable.

Validity of the knowledge test was indicated through content validity. This process ensures that the items of the knowledge test are drawn from the domain of objectives set out in the module. Each objective is represented by one item in the knowledge test. The test items were selected to ensure that no aspect of the unit was over represented in the tests.

### **PROCEDURE**

The teachers in the experimental group were intensively inserviced on the methodology of teaching that was required to ensure uniformity of treatment in the four experimental classes. This inservicing explained that the treatment to be given to the experimental group was to involve the use of formative and summative assessment. The treatment would involve revising previous lessons, setting homework and home study. Students would be quizzed on work covered during the module, given feedback in each subsequent lesson and frequently motivated to prepare thoroughly for the final test. Normally this approach to teaching has not been part of the methodology

of teaching religious education in Catholic schools in Western Australia. The control groups would not receive this treatment nor would the teachers in the control group have this information. Observation and recording of teaching in the control group is used to confirm the level of use of systematic assessment procedures.

Each teacher in the experimental group was given a teaching program and daily lesson plans. The lesson plan included review questions, homework and class work. In an effort to prevent teachers teaching to the tests, none of the teachers had access to test papers until the morning designated for each particular test. The daily review tests were administered to the experimental group, collected and marked by the researcher, and returned prior to the next lesson. The teachers then went through each item, corrected any misunderstandings and directed students to correct errors or incomplete answers. All classes were given a pre-test prior to the commencement of the study. All classes were given the same test as a post-test at the end of the four week module. Two weeks later, after two weeks of holidays, a follow-up test was administered.

### **RESULTS**

Table I summarises the scores of the knowledge tests. The knowledge pre-test scores indicate that no one class has a score in the knowledge pre-test that is markedly different from any other class. The mean score on the knowledge pre-test for each class also

indicated that no significant knowledge of the content of the unit existed. The sample mean was 5.14.

Each individual class had similar results with a similar distribution. The mean scores of the experimental and control groups were also very similar: 5.22 and 5.05 respectively.

Class	Pre-test Score	Mean	Post-test Score	Mean	Follow-up Test Mean Score
1	6.0		13.5		12.2
2	5.2		9.9		9.3
3	4.9		12.6		10.7
4	4.8		11.2		10.0
5	4.7		5.9		5.4
6	5.0		5.6		5.2
7	5.3		4.9		5.6

Classes 1 - 4 are the Experimental Classes; Classes 5 - 7 are the Control Classes.

The results of the post-test illustrate that a difference exists between the experimental and control classes. Each of the experimental classes scored mean post-test results well above the means of the control classes. The experimental classes had means of 13.5, 9.9, 12.6 and 11.2 while the three control classes had mean scores of 5.9, 5.6 and 4.9. The mean score for the experimental group was 11.9 while the mean score for the control group was 5.4.

Each of the four experimental classes had scores that improved after the pre-test. The small amount of change in the scores of the control classes is also very evident.

The mean score for the experimental classes was 10.65, with the mean scores of the four experimental classes ranging from 9.33 to

The difference between the experimental and the control groups, when the knowledge pre-test scores are considered, is not significant at the 0.05 level. An ANOVA of the results of the seven classes indicates that no two classes are significantly different at the 0.05.

12.22. The mean score for the control classes was 5.44.

The mean scores of the control classes are little different from the pre-test scores. The ANOVA shows that there is no significant difference, at the 0.05 level, between the pre-test and post-test scores. The level of change for the four experimental classes was more substantial. This result indicates a significant level of difference in knowledge learning outcomes at the two levels of the nested design. The nested design analysis indicates that the variation in post-test knowledge scores is significantly different at the 0.05 level when "method" is considered. The results of the analysis indicates that differences between individual teachers was not significant at the 0.05 level.

## **DISCUSSION AND IMPLICATION**

Significant differences between the control and the experimental groups were evident when the knowledge test scores were analysed. The pre-test scores indicated that all students had similar levels of knowledge prior to teaching of the module. At the post-test stage the control group had shown little change in test scores while the experimental group illustrated significant change in test scores. The difference between the two groups was maintained at the follow-up test stage.

The results clearly supported Rossiter's view (198 1) that a relationship exists between clarity of purpose and learning outcomes. The students who received the "treatment" had results that were significantly better than the students whose teachers had not been exposed to the same specific directions. The nested design of this study allowed individual classes to be compared, as well as a comparison of the experimental group and the control group. In both instances, the results of the experimental classes were significantly different to the results of the control classes. The results of each control class were similar, and indicated that no learning of content had occurred. The results of the four experimental classes were similar to each other and indicated a significant positive change in knowledge test scores between the pretest and the post-test. These results therefore seem to support Rossiter's contention that clarity of purpose can directly influence learning outcomes. The four experimental classes had direction and

purpose. The control classes did not have this level of clarity.

Content that had been covered by the teachers of the three control classes seemed to have not been learnt. Post-test and follow-up knowledge test results indicated almost no change in knowledge test scores from the scores attained by the students prior to the module of work beginning. Teachers were "teaching" but the module content was not being learnt. This result was in contrast to the observed outcomes of the experimental classes. Here, teachers imposed a formal assessment structure, actively revised each lesson, set minor tests, reviewed material and actively utilised many forms of formative assessment. In these classes students learnt the material that was being taught. Knowledge post-test scores were significantly higher than the pre-test scores. Learning was shown to be long term as the follow-up test results were also significantly higher than the pre-test scores.

The differences between the experimental and the control post-test and follow-up knowledge test scores cannot be explained by differences that existed between the classes prior to the study beginning. This has been shown with the analysis of pre-test data. With each set of results there were no significant differences between classes for any of the indicator variables. These included indicators of religious background, commitment to religion, home study, prior knowledge of the unit of work and reading ability. The observed

differences in knowledge test scores must therefore be associated with the treatment the experimental classes received during the study.

In calculating the within-group difference as well as the between-group differences, the nested design analysis allows comment on the possible differences between each teacher in the control and experimental classes. While every care was taken in the experimental design to randomly allocate teachers to each class, some advantage could have occurred for the experimental classes. These teachers may have been more dynamic, more committed and more inspirational. The nested design analysis indicated that when the scores of individual classes were compared, there were no significant differences. This pattern was evident for the knowledge test at the pre-test, post-test and follow-up test stages. The nested design analysis indicated that there was no significant difference between any of the four experimental classes when the post-test and follow-up test results were considered. Similarly, the analysis indicated that there was also no significant difference between any of the three control classes. This indicates that teacher differences in this study did not significantly influence the knowledge test scores. It would seem that the difference in test scores was the result of the difference in teaching.

A significant theme in the literature pointed to the effect of poor teaching within religious education in Catholic schools. This perception

was shown to hold true within the study school. Observation of the control group of classes indicated that the teaching lacked academic rigour. No tests were planned: teachers failed to utilise any structured formative or summative assessment procedures. In these three classes knowledge test scores were very low. Scores at the end of a four-week module were barely different from the scores recorded in the pre-test. No learning appeared to have taken place. On the other hand, the four experimental classes showed significant changes in knowledge test scores. Teaching in these classes included systematic formative and summative assessment. They were shown to do much more study. It would seem that the concern expressed in the literature regarding teaching technique in religious education is supported by the results of this study.

The problems facing religious education in Catholic schools have been viewed too exclusively as problems of "religion" rather than problems of education. The literature faces this issue from an educational perspective. The literature calls for a more professional approach to the teaching of religious education. This professional approach involves determining objectives, determining classroom process and designing methods for determining whether the classroom processes achieved the objectives. Thus the need for assessment and evaluation is integral to good education. As good education is integral to religious education, the inclusion



of assessment and evaluation is crucial for a professional approach to teaching religious education in Catholic schools. The results of this study confirm that the use of assessment and evaluation in the teaching of religious education is of benefit to both the student and the teacher.

The "treatment" experienced by the students in the experimental classes did not focus solely on measurement. All aspects of assessment/evaluation were utilised. Daily tests provided immediate feedback to the student regarding both poorly- and well-learned material. Poorly-learned material could therefore be revised by the student. Well-learned material could be built upon in subsequent lessons. The teacher received immediate feedback on aspects of teaching that were inadequate and hence needing improvement. Interviews with students confirmed this potential benefit. Students who received the "treatment" recognised the potential benefit. They saw that testing gave them feedback

that could be acted upon and motivated them to improve their learning. These benefits obviously flowed through to post-test and follow-up test scores. The students in the control classes did not receive the treatment. Taping of lessons revealed a low level of formative and diagnostic evaluation which tended to be very lesson specific. Evaluation did not extend to previous lessons nor act as a motivator for learning. The results of this lack of "treatment" flowed through to the post-test

and follow-up tests which indicated that very little learning had occurred.

In addition to the improvements in teaching and learning, the use of assessment procedures has a by-product effect. It is suggested that students may perceive religious education in Catholic schools as having little significance. Literature in the area of religious education supports the observation that the poor status of the subject in Catholic schools is related to the subject being non-examinable. The results of this analysis support this view.

The students in the classes who were told about the final test performed at a significantly higher level than those who had no knowledge of this end-of-module test. The focus of this long term goal was maintained with daily tests. Students knew that each day their learning would be tested and their results constantly reviewed. Students quickly see the direct connection between the effectiveness of their home study and the results of their daily tests.

Interviews with the students confirmed the connection between assessment and status. Many students reported that they felt "good" about their test results. They reported that they felt that religious education was "more interesting", "valuable" and "worthwhile". They felt that they were achieving something each day. These reports were in contrast to the students representing the control group. In addition to benefits to teaching and learning, student perception of religious education can improve where assessment and evaluation

procedures are utilised as part of the teaching methodology.

### **CONCLUSIONS**

At level one of the nested study, a clear difference between the experimental group and the control group is observed. This difference was evident not only at the post-test stage but continued beyond the teaching phase and was evident in the follow-up test. These results indicate that the treatment was able to produce significant change in knowledge learning outcomes. The treatment involved the use of assessment and evaluation procedures in the teaching of religious education. The control group was not exposed to this method of teaching. The results of the control group indicated that no significant change in knowledge learning outcomes occurred between the pre-test, post-test and at the follow-up test stage.

Analysis of a range of indicator variables which might have an influence on student learning indicated that there was no significant difference between the profile of the control and the experimental groups. Relating knowledge test scores to these variables indicated no significant relationship. Knowledge test scores did not significantly vary when each factor was considered. A student's religious background and commitment to the Catholic religion did not appear to impact on knowledge learning outcomes.

The elimination of each of these extraneous variables leaves the "treatment" as an intervening variable on student learning outcomes. The differences in knowledge learning outcomes can therefore only be accounted for by the difference in teaching methodology.

At level two of this nested design the conclusions are the same. Level two considered individual class differences. The analysis of knowledge results indicated that while small differences in knowledge scores were evident between each of the four experimental classes, these differences were not significant. This was the case at all three stages of testing. The same outcome arose when the knowledge scores of the three control classes were compared. Individual teacher differences therefore did not complicate student learning outcomes in

this study.

Each of the four experimental classes scored significantly higher knowledge test results than each of the three control classes. The extraneous variables (religious background, commitment to the Catholic religion) were also considered at level two of this analysis. No differences were evident, indicating that all classes had similar personal and family characteristics. These factors were shown to not have any significant effect on student learning outcomes.

The results of the study are clear. The use of a more academic mode of teaching, with its

associated assessment and evaluation procedures, in religious education in Catholic schools does affect the knowledge learning outcomes of students. The learning effect is significant and positive. The students who did not receive the treatment indicated little change of knowledge scores. The students who did receive the treatment demonstrated significant gain in knowledge scores. Therefore change in knowledge scores was not the result of other factors but may be directly attributable to the teaching process.

Some teachers of religious education believe that their subject is different from subjects such as mathematics, science and history. They believe they can teach effectively without the benefits of assessment and evaluation. It is important to consider the results of this study in the light of incorporating assessment and evaluation procedures in the teaching methodology of religious education.

### **IMPLICATIONS**

This study clearly indicates an improvement in test results after the introduction of assessment and evaluation strategies. While such an improvement would be not unexpected by most educationalists, the resistance of teachers of religious education to the utilisation of assessment strategies, suggests that these teachers may well not expect the observed results. Researchers in the field of religious education developed a sound philosophical framework that supported the utilisation of assessment strategies in the teaching of religious education. This study implemented

the philosophical framework into an actual classroom situation.

Two significant implications arise. Firstly, the test results of the students in the control group indicate that they were not actually learning any of the objectives of the course. The teachers who refuse to utilise assessment strategies in their teaching are not achieving the objectives of the course. Without the assessments strategies these teachers do not know that they are not successfully teaching the students; they do not know if the teaching strategy they select is the best strategy for those students.

The teachers who did utilise assessment strategies do have information about their teaching and how well their students are learning. They can make informed judgements about their selected teaching strategy and adjust the strategy as the teaching progresses. These teachers are in control of their teaching.

The results of this study also seem to indicate that the use of assessment strategies may have a motivating effect on students. The knowledge that the objectives will be assessed may well spur students on to greater effort in their class-work, study and revision. Students who know that their is no summative assessment may well feel that their is little need to study and revise the objectives of the course.

The implications to teachers of religious education are clear. Effective teaching and

learning require an effective assessment and evaluation strategy.

**REFERENCES**

Cole, P.G. and Chan, L.K.S. (1987). *Teaching principles and practice*. Sydney: Prentice Hall

Crawford, M.L. and Rossiter, G. (19 89, February). Religion teachers as missionaries to a teenage culture. *Word in Life*, pp. 17-20.

Kubiszyn, T and Borich, G. (1987). *Educational testing and measurement. Classroom application andpractice*. Second Edition, Glenview, Illinois, London. Scott, Foresman and Company

Rossiter, G. (1981). The gap between aims and practice in religious education. *Word in Life*, 29, 24-28.