2001

Generic versus content-driven assessment

Jim S. Tognolini

University of Wollongong, jimto@uow.edu.au

Generic versus content-driven assessment

Abstract

In my Masters class I pose the question "How would you respond to the statement : Schools should teach students skills rather than content"? In replying to the question I expect students to make the point that while skills are important, they have to be applied to something, that something is content. I firmly believe that the dichotomy set up between content and skills is a false one. The relationship is one of emphasis and relativity. This paper considers how some of the states and territories have addressed or are currently addressing the relationship between skills and content in a number of their assessment activities. In addition, it examines the idea of authentic assessment and posits an argument that it might be time to challenge the rituals associated with content-based and generic skills-based examinations.

Keywords

generic, driven, assessment, content, versus

Disciplines

Education | Social and Behavioral Sciences

Publication Details

Generic versus Content-Driven Assessment

A Paper presented by

Professor Jim Tognolini

Educational Testing Centre, UNSW

at the

ACACA

(Australasian Curriculum, Assessment and Certification Authorities)

Conference 2001

at The Grace Hotel, 77 York Street, Sydney

on 26 July 2001

ABSTRACT

In my Masters class I pose the question "How would you respond to the statement : Schools should teach students skills rather than content"? In replying to the question I expect students to make the point that while skills are important, they have to be applied to something, that something is content. I firmly believe that the dichotomy set up between content and skills is a false one. The relationship is one of emphasis and relativity.

This paper considers how some of the states and territories have addressed or are currently addressing the relationship between skills and content in a number of their assessment activities. In addition, it examines the idea of authentic assessment and posits an argument that it might be time to challenge the rituals associated with content-based and generic skills-based examinations.

Generic versus Content-Driven Assessment

When John asked me to talk about the topic *Generic versus Content Driven Assessment* I was at a bit of a loss as to what I could speak about. John's prompts regarding "Authentic" and "Performance" based assessments did not give me any immediate clues either. However, in my naivety I agreed to speak on the topic, thinking that the 26th July was a long way off and I would be sure to get some inspiration before then, so everything would be alright. It is now the middle of July, the 26th July is perilously close and I am still grappling for a thread to link the topic together.

When I first started at the Secondary Education Authority in Western Australia I attended an Australian Curriculum, Assessment and Certification Authorities (ACACA) meeting in Sydney. I remember coming back to the hotel where we were staying in Bronte deeply engrossed in a discussion with John Pitman. I further remember John arguing very strongly that content-based examinations were basically invalid and that generic assessments of the type encapsulated in the Australian Scholastic Aptitude Test (ASAT), supplemented by a writing task were a much more valid way of assessing student learning. He argued that it was consistent with a theory of learning, which espouses that students learn best when they can perceive the relevance of learning and the assessment activities enhance learning outcomes. At the time I remember arguing vehemently with John that content-based examinations were much better than the generic skills based tests. I cannot remember why I argued for this other than that is what we did in Western Australia.

Over the years John and I have reflected on the topic a number of times in a number of different locations. The penultimate time being when he, Mike Steer and I discussed it near the camel corral of a Bedouin camp outside of Jerusalem at the 2000 IAEA Conference.

Anyway I thought I would take the opportunity to visit John in Queensland and ask him what he thought I should talk about at this Conference. What became obvious to me as we talked over lunch was that either we had both mellowed over the 20 or so years that we have been friends and colleagues or that we had not been listening to each other in those earlier discussions. Because I felt that we were in fact saying much the same thing and it is what Susan has just reiterated in her presentation: … the topic generic versus content-driven assessment establishes a false dichotomy – all systems have both generic and content-based assessments in gauging student learning.

I eventually came to the conclusion after talking with John that I would take the opportunity to discuss a number of disparate, but interrelated issues involving generic and content driven assessment.

The main message of the paper is that all systems are interested in teaching and assessing generic skills. While most teach the skills within subject boundaries they acknowledge the value in students being able to generalize these skills across subject boundaries. This paper shows some of the ways systems have firstly, assessed the skills within subjects and then, how others have tried to use formal assessments and curriculum developments to encourage schools to teach students how to generalize the skills across the boundaries.

Generic Skills and Curricula

As some of you may know the Educational Testing Centre (ETC) conducts school based competitions in all states and terrtories in Australia and in a further 33 countries overseas. In a large number of those countries we cannot even read the curriculum documents because they are in foreign languages. How can I be sure that we are testing elements of

the curricula with our competitions? The answer is simple, because we construct items to measure the generic skills within the context of science, mathematics, English and computer studies. That is we write items to measure such skills as problem solving, critical thinking, reasoning, comprehending, interpreting information and communicating in the various subject areas and we know while the content may be different from state to state or country to country, these skills will be embedded in one form or another in the curriculum documents.

The same principle applies to the Scholastic Aptitude Test (SAT) used by many US universities as the competitive medium for entry for prospective students from within the US and around the world. In the US an instrument such as the SAT has to be used because there is no established national or even state syllabus. In fact within a state you could have anything up to seventeen hundred school districts all following their own syllabi and constructing an examination based upon content within such syllabi is impractical.

The SAT is a test that effectively assesses student performance on generic skills while controlling for the different content that the students may use to develop those skills. It is a common test that enables students to be compared on a common scale even though they may have done different courses within the broad based disciplines of mathematics, science, the humanities and the social sciences.

The Australian Scholastic Aptitude Test constructed by the Australian Council for Educational Research (ACER) is a similar type of test. It used to be used each year by Western Australia, Queensland and the Australian Capital Territory as part of their scaling activity for constructing Tertiary Entrance Scores (TES). In Western Australia we used it to eliminate differences between subjects. That is we used it to ensure that students were neither advantaged nor disadvantaged by the subjects that they used to create their Tertiary Entrance Score.

Statistical moderation using information from content-based examinations, on the other hand, was used in Western Australia to ensure that students were neither advantaged nor disadvantaged by the school-based metric used to create the school-based assessment. In the ACT and Queensland, where there were no external examinations the ASAT was required to take account of both differences between subjects and schools. As a consequence we spent a lot of our time at the users group looking for any systematic effects that could influence the scaling activity. This came to a head in the eighties when it was postulated that there was gender bias in using ASAT because of the differential performance of boys and girls on multiple-choice items. While this was not considered to be a big issue in Western Australia, where the unit of analysis was the subject cohort it was a huge issue in both the ACT and Queensland where the unit of analysis was the school and some single-sex schools were perceived to be systematically disadvantaged by the use of the scaling test.

The other states did not use ASAT as the common test to bring the scores from the different subjects onto the same scale so they did not have to worry about the issue.

The key point about this reflection is that all states and territories have generic skills embedded within the various subject curricula and therefore there has always been a need to ensure that the skills are assessed. Differences amongst states and territories manifest themselves in the way that the states and territories encourage schools to teach and assess these skills across subject boundaries.

Assessing Generic Skills Across Subject Boundaries

By their very nature the types of skills alluded to earlier are not subject or content specific. They extend across the subject areas. However, because of the nature of secondary schools where students are taught subjects within Key Learning Areas, and hence skills have been developed within these subjects, there has been a tendency for students and teachers to associate the skills with a subject rather than be able to generalize the skill across subjects.

How many of us who have taught middle school students have had the experience of students having been taught, for example, bar charts in mathematics, going to a geography class and telling their teacher that they have never seen a bar chart.

There has been a hundred years of psychological research showing that students have great difficulty generalizing their skills across subject boundaries and this has been a source of contention for educators around the world. In Australia we have tried a number of techniques to ensure that we are not only assessing skills within a subject domain, but also assessing them across subject boundaries (i.e. across the curriculum). In addition, we have tried to raise the consciousness of teachers as to how important it is for student learning to have children apply skills across subject boundaries. While children will learn this in time, they can learn it quicker if teachers encourage them to do it.

I will now reflect upon some of the ways that the different states have tried to assess generic skills across the curriculum and by doing so tried to send a message to teachers that it is their responsibility to not only develop and nurture skills within their subject, but also to make their students aware of the fact that the skills are generalizable.

- **Western Australia**

In Western Australia in the mid-eighties we considered how we might assess literacy and numeracy across the curriculum. The method that we eventually invoked involved ensuring that literacy objectives were included in each subject's syllabus. Teachers from all the different subjects were then required to prepare a portfolio of student achievement on the various literacy and numeracy outcomes within their subjects. They had to keep records showing the most recent piece of work testifying to the achievement of each student in their class relative to the literacy and numeracy outcomes in each subject that they taught. You can imagine the logistical nightmare associated with carrying out such an activity. I remember one school putting in a request for 15 extra filing cabinets to store the data. One of the problems with this method was that the only person who really had any concept of the generalisability of the skills across the curriculum was the form teacher who had to compile the different reports for his/her students across the subjects. Otherwise the teachers still taught the skills within the confines of their own subjects and reported the results accordingly.

This method did not last long. However, the idea of teachers assessing student performance on say computing skills, problem solving skills and some affective skills within their subject areas and then having this information aggregated for each student, by teachers, across subject areas is still being considered as a means of assessing generic skills.

The challenge is still there, we need to not only assess generic skills validly within subjects but also help children be able to learn how to apply these skills across subjects.

We are all conscious of the power of 'high stakes' assessment to influence teacher behaviour. In a recent On-line paper by Allan Luke (1999), he makes this point quite forcefully

"… Thirty years of research on teacher behaviour tells us that as soon as we bring in 'high stakes' assessment, teachers will suss out what is being assessed and reorient the curriculum and pedagogy back towards 'high stakes'.

The point was also brought to my attention just last week when I was in Singapore. The curriculum in Singapore has been augmented by a requirement that students be taught how to think. Teachers have had training in how to teach their children to think, both implicitly and explicitly. While I believe the schools have taken up the challenge, an instruction that has been sent around one of the most prestigious schools is, *"We will stop thinking six weeks before the examination and focus entirely upon preparing for the examination".*

Given the power of 'high stakes' assessment to drive curriculum, it is critical that there be a synergy between the 'high stakes' assessment of a state or examination authority and the desire to get teachers to help students generalize across subjects. Queensland is one system that has used 'high stakes' assessment at Year 12 to encourage teachers to operate outside their subject specific 'silos'.

- **Queensland**

In the early nineties, the Queensland Board of Senior Secondary School Studies examined syllabi from Senior Board subjects and identified the curriculum elements in the different syllabuses. They then identified, through a continual process of refinement, the common elements across the syllabuses. These elements were then assessed using items within the context of 5 broad disciplines – humanities, mathematics, physical sciences, applied sciences and social sciences.

The items in the tests (Core Skills Tests) assess the common elements and the content that provides the context for the items is provided within each item.

By focussing attention on the common elements (generic skills) using a relatively high stakes testing situation, while at the same time both broadening the context for the item and taking away the requirement for content specific knowledge by providing it in the item, schools could be encouraged to not only focus upon assessing the skills within the subject, but also assessing the skills across a range of subjects within broad based discipline categories (or fields of knowledge).

John Pitman made the point that a number of schools in Queensland were starting to break down the "subject silos" by providing tasks to students that focussed upon the skills and required students to solve the problems by drawing upon the content as and when it is required.

Of course, the potential for this cross-subject assessment of skills exists irrespective of the existence of skills-based tests. While the examinations for those states that have examinations are undoubtedly content-based, all states also include a school assessment component that enables schools to assess those components of the curricula that don't lend themselves to examinations. In schools, it is possible to assess skills (that are now contained in outcome statements) across a range of areas using such assessment devices as research projects, investigations, oral presentations and practical tasks and includes the information in the final assessment of the students. However, it must be pointed out that the incentives for the schools to do this in systems that contain moderation by school-based assessment are reduced because of the nature of statistical moderation. As a consequence it could be argued that content-based assessments inhibit the evaluation of generic skills outside the specific subject.

In saying this I am not saying that content-based examinations do not assess generic skills. I will take this issue up later in the presentation.

The Victorian GAT is another test focussing upon generic skills that is used as part of the VCE battery of examinations and sends the message to teachers that the teaching and assessment of these generic skills is an important part of the students' learning.

- **South Australia**

South Australia has recently introduced a new framework called the South Australian Curriculum, Standards and Accountability (SACSA) Framework. This Framework draws on a constructivist view of learning and explicitly identifies five Essential Learnings, together with concepts and processes drawn from the Learning Areas to provide the connecting threads for the whole curriculum. The Essential Learnings are: Futures, Identity, Interdependence, Thinking and Communication.

Having a constructivist base means that there is an emphasis on holism and on students making connections and this in turn sets up the potential for developing learning across subject areas as well as between stages of education.

Insert Figure 1 Here: SACSA Framework Model

At year 12, the outcomes and standards are described in terms of the Essential Learnings capabilities and the outcomes described by external curriculum sources. As far as I can ascertain the Essential Learning performance of students will be monitored by teachers using a range of assessment devices and the result will be referenced to articulated standards. The challenge, as is the challenge, for all generic skills that require assessment across the curriculum, is to produce a composite image of a student without just

aggregating the component parts obtained from within subjects. Perhaps in the first instance the Victorian example where different teachers do assess students on the generic skill and a composite picture is then reported might be the way to go. However, using assessment tasks that transcend subject boundaries to assess these learnings is the most appropriate method for assessing them. But there are a number of questions to be addressed before this can be done at the upper secondary level: What would the tasks look like? How would schools accommodate the cross curricula interaction in their current structures? Can the teachers of English, mathematics and Design for example, combine to not only set an integrated task to assess an aspect of Futures, but also agree on the marking key? What training would teachers need to be able to do this effectively? How do you get objectivity in assessment? How do you get comparability between results?

Questions similar to these (but referenced to other generic skills) are being currently addressed in the 'New Basics' Project in Queensland. This project was initiated by Allan Luke and is now directed by Gabrielle Matters. In essence, the project involves developing rich tasks that are drawn from (emergent and residual) forms of life, rather than fields of knowledge. The fields of knowledge (KLAs) will be used to service the tasks rather than vice-versa. The tasks are intended to be inductive rather than deductive, holistic rather than positivist. Teachers begin with the whole task that has visible value in terms of 'real world' relevance and use their knowledge to distil down knowledges, sub-skills and competencies for teaching (Luke, 1999). The work that is being undertaken in this project will contribute significantly to the understandings of how generic assessment might 'drive' the curriculum.

Generic Skills Within Content-driven Assessments

I do believe that a broad range of generic skills within subject specific boundaries are assessed quite effectively in subject –based examinations This was demonstrated to me

quite recently in a project I was involved in at the NSW Board of Studies. The Project involved developing Performance (Achievement) Scales for Higher School Certificate (HSC) subjects.

Two approaches were used in developing student performance scale descriptions (standards). The first involved "experts" developing descriptions of bands of student achievement based on expected and realised performance from the 1996 HSC examinations. That is the experts used a modified Angoff approach to establish cut-scores for each band of performance. They then had student scripts representing performance at these cut-scores presented to them and they used these to describe the typical performance of students in these bands.

The second approach involved analysing the 1997 HSC using the Extended Logistic Model (Rasch Model for polytomously scored data). The item descriptions, when located on a scale, provided content-based descriptions of what it is students can do at different locations on the scale. Expert teams comprising subject specialists used this information in conjunction with this intimate knowledge of the course to write descriptions of student performance, which captured the six levels (Band 1 to Band 6) comprising Stage 6.

After each expert team had completed the task for all subjects, it became evident that there were similarities amongst similar bands across different subjects.

The Board of Studies considered it useful therefore to try and standardize some aspects of band description statements across the subjects. Consequently it set out to create a Generic Performance Scale that would then be the template for subjects to modify their draft performance scales.

The process involved

1. collating all performance scale band statements across subjects within bands;
2. identifying generic "skill categories" to represent like clusters of the statements (initially 17 categories); and,
3. summarising the statements into "skill categories".

Initially 17 categories were identified. However, after consultation this set was further collapsed into 8 categories or dimensions:

- Knowledge and Understanding
- Practical Skills
- Research Skills
- Problem Solving Skills
- Communication Skills
- Originality and Creativity Skills
- Inference Skills
- Analytical and Evaluative Skills

The original performance standards were then re-evaluated for the various subjects to ensure that, where appropriate, the 8 dimensions were represented and the descriptive language used was relatively consistent in the performance standards for the subject and are described as part of the draft syllabi.

The exercise was designed to standardize to some extent the nature of the band descriptors across subjects. However, as part of the exercise the skill statements were compared to the descriptor statements from the NSW Key Descriptor Project and the Queensland Core Skills Test Common Curriculum Elements and it was found that they

were consistent with both. The broad categories were also consistent with the Mayer Key Competencies and to some extent Blooms Taxonomy.

An interesting point about this analysis is that we arrive at basically the same set of skills from a number of different analyses. In the Queensland Project the skills were derived from an analysis of the curriculum, rather than an analysis of the test acting as the curriculum.

Of course the number and range of the skills in the Queensland Project is much greater than that of the NSW Project because the former identified all the skills in the syllabus documents where the latter only identified those that are assessable by external examinations. If a similar analysis had been conducted using skills assessed by school-based assessment the lists would probably have been the same. The point is that generic skills are assessed in subject-based examinations just as they are assessed in tests that are designed specifically to assess generic skills.

A second point is that because the items in the Core Skills Tests are assessing the same skills as the subject-based HSC examination (albeit with different item types), then, in theory anyway, the image of individual students formed by performance on the tests should be relatively the same, irrespective of whether they are assessed by syllabus-based examinations or skills based tests. That is there should be a high correlation between performances of students on both measures. We know, however, from correlation studies done between student scores on ASAT and GAT tests and their scores on subject-based tests that the correlations are quite low. The question is why? One of the reasons is that while the skills being assessed might be the same the context for the items is quite different. In tests that assess generic skills there is generally a requirement that the content required to answer the item is provided in the item. Whereas in the subject-based tests the content is generally asked as part of the question and the skill component requires the student to know and use this content in solving the item. This means that

although the items are measuring the same skill they are really measuring slightly different things.

Insert Item from Year 12 Science Competition here to demonstrate the use of content usage

The key point I am trying to make here is that generic skills are assessed in examinations like the HSC and that these skills are the same as those assessed in tests and tasks that are specifically designed to assess generic skills. The difference hinges on the breadth of the context used in assessing the skill and whether or not critical content is supplied as part of the item or is generally assumed knowledge.

A related point is that sometimes we get so involved in our own learning theories that we forget to give credit to the students. While I am cognizant of my previous example regarding the ability of students to generalize skills across subjects, I am reminded of an example in China when I was talking to a group of teachers about testing. They were telling me that an item assessing vocab in context would not be able to be done by their students because the word 'astronaut' was not one of the 800 English words that they had been taught up until Year 9. I encouraged them to give the item to their classes. They did so and were pleasantly surprised that over 90% of the students answered the item correctly. Of course they shouldn't have been surprised because students learn from a variety of sources (e.g. television, friends, family, books). Similarly, as students get older the skills they learn and develop within subjects do get generalized, as they have to solve problems, reason and communicate in their lives. **Ultimately I wonder just how different the final image of the same student would be if he/she were trained and assessed in two different systems.**

Maybe John Pitman and I were both right 20 years ago and that's why we seem to have mellowed over time.

Authentic Tasks

It would seem appropriate to talk a little bit about 'authentic tasks' for a number of reasons. Firstly, John Ward suggested that I should do so when he first talked to me about giving this presentation. Secondly, there is a strong demand to ensure that, where possible, assessment of learning should be contextualised, and meaningful to students. This equates on this occasion to trying to make the problems more 'real world' like. Thirdly, because my co-presenter indicated that I would address some of the more measurement oriented issues associated with the topic.

There are a number of different aspects of '*authentic assessment*' that could be addressed. Cummings and Maxwell (1999) give a good overview of some of these while also linking authentic assessment to performance-based assessment, problem-based assessment and competence-based assessment. I want to just concentrate on one small aspect of test construction; namely the relationship between authenticity and 'real world'.

It seems as though testing agencies are being asked to produce items that simulate real life situations as part of their assessments. Hence test constructors go to great lengths to simulate 'real world' situations into their tasks. I want to consider one example that shows some of the difficulties that arise from the lack of clarity that can emerge when mathematical rigour and examination objectivity for example, interact with everyday discourse.

Insert Mathematics Item about the Lift here

I have chosen an item designed to test the outcome '*solve number problems with the aid of a calculator*'. The answer is 269 divided by 14 (i.e. 20). This indicates that they have interpreted the display to select the most appropriate whole number in this context.

The question is, "Is the answer 269 divided by 14 rounded to the next whole number, correct?" Well possibly, if the student realises that it should not be treated as a real 'real life' problem. For example, the child has to assume, at least implicitly, in order to get the correct answer

- that there are always 14 people in the lift (i.e. it's always full);
- that no-one, particularly those on the first floor, in the morning rush for the lift gets sick of waiting and decides to use the stairs;
- that everyone uses the normal space associated with a person in a lift. That is, no one uses a wheel chair.

In other words, the child has to ignore the reference to 'real world' to solve the problem. One of the dangers with such problems, presented and marked in this way, is that the more able students are likely to be confused by such items and hence the item will not function effectively – it is also manifestly unfair. That is why I indicated earlier that the task of constructing tasks for the 'New Basics' would be challenging, particularly if there were a need to compare results. This would require the administration and marking of the tasks to be standardized.

As Cummings and Maxwell (1999) state " an important characteristic of simulated tasks is that they are not actually real". Swanson (1998) also makes the point slightly differently, as follows

> " **No matter how realistic a performance-based assessment is, it is still a simulation, and examinees do not behave in the same way they would in real life. Neither traditional testing nor performance-based assessment methods are a panacea. Selection of assessment methods should depend on the skills to be assessed and generally, use of a blend of methods is desirable**"

This does not mean that we should not strive to produce tasks that are 'real world'. It does mean that we are faced with a challenge. How can we move towards authenticity, particularly in 'high stakes' tasks that require standardization in terms of administration and marking?

A Final Challenge

I would like to throw down my final challenge. It is important to be continually reviewing the way that 'high stakes' assessments are supporting the learning theories that underpin curricula. When conducting the review I would suggest challenging the rituals that currently surround external assessment. For example, would the examination be valid if students were to bring books and notes into it? Should the examination be a three-hour examination, or if it is three-hours, should typical students be able to complete it in 2 hours, thus eliminating the speed component in typical examinations? Should there be a performance component effectively incorporated into an examination? Should the examination be problem-based with the students being able to access information from a range of sources during the examination? Can the examination be carried out on a computer in which students have unlimited Internet access? If we really want to simulate authenticity then maybe this last question is one that should be considered first. However, before it can be done, the nature of the syllabuses would have to be reviewed and teaching pedagogy would also have to be reviewed. By the way, in my opinion it is not a case of if, it is a case of when will we move to this as a means of assessing student learning.

Conclusion

There is no doubt that all systems are seeking to assist students to nurture, develop and generalize their skills across subject boundaries. Some systems are being more proactive than others in encouraging teachers and systems to move outside their subject boundaries when applying their skills. Some use 'high stakes' assessment, others use curriculum development, while still others will move towards this as they take-up outcomes-based assessment. I am not convinced that the end products of these systems would be drastically different. However, I do not know whether this is the case. I do know that to change teachers and make them function outside the security of their subject blankets is a huge undertaking.

I think some key issues for discussion might include the following

1. How do you assess generic skills across boundaries when schools are operationally oriented to teach them within subjects?
2. Is it possible to get secondary teachers, steeped in the traditions of their subjects, to teach students to generalize skills across subjects? How?
3. What type of assessment tasks can be used at Year 12 with 'high stakes' assessments to effectively assess generic skills across subjects?
4. Is it adequate to get teachers to assess the generic skills within subjects, report these, have someone collate the reports and then make an "on-balanced-judgment" about the student's performance on the skill or set of skills being assessed? What are the weaknesses of this type of assessment?
5. Should we strive to engender authenticity into our assessment tasks?
6. Is problem-based learning a possibility in the near future? What might public examinations look like within the next five years?

Finally, I come back to my original point. The topic *generic versus content-driven assessment* establishes a false dichotomy – all systems use both generic and content-based assessments in gauging learning; it is a case of different emphasis. Some emphasize the assessment of generic skills across subject boundaries more explicitly than others. If I had a crystal ball I would see, in ten years time for example, different types of curricula and examinations across all systems. Students will have access to technology that will mean that knowledge is readily available to them. The examinations will be testing the generic skills and there will probably be no subject boundaries. Who knows?