Assessment for Learning: Using SOLO Taxonomy to Measure Design Performance of Design & Technology Students

C. F. LEUNG

Block M, Flat 704, Kornhill, Quarry Bay, Hong Kong, P.R. of China

ABSTRACT: One objective of Design & Technology (D&T) is to enable students to be inventive in designing practical solutions to problems. D&T is viewed as being successful when students can demonstrate the ability to recognize problems, undertake inquiries by themselves, and contribute ideas accordingly. This article will discuss a study which investigated an alternative approach to assessing students' design performances. In the study, a new item format was designed and a new criterion framework of assessment based on Biggs' SOLO Taxonomy was developed. The evidence from this study indicates strong face validity for the new approach which maps closely to the goals and purposes of learning D&T.

Keywords: criterion-referenced assessment, design & technology, design performance, educational assessment, SOLO taxonomy

LEARNING AND ASSESSMENT

Assessment incorporates emerging ideas in the understanding of learning, including the construction of meaning, the importance of prior knowledge, and the strategies for representing knowledge. This thought leads to new criteria of assessment validity which are different from the traditional psychometric perspective. The emphasis of the test on meaningfulness supports the motivation and sense of utility that students and teachers perceive. There must be evidence that the measures respond to differences in instructional intensity, academic exposure, and other indicators of the quality of educational processes. The fundamental purpose of assessment in schools shifts to that of monitoring the learning progress of all students against progressive standards and using the results to inform the teaching and learning process.

Systematic assessment can provide useful information on learning outcomes. Aligning with a trend to improve the quality of education, the assessment movement has become more standards-based and more learning-orientated. To enable effective teaching and learning, criterion-referencing assessment showing evidence and recognition of learning progress is an important motivating factor and an incentive to further learning. Assessment is an integral part of effective learning, whereby students are provided with comments on progress. Where assessment is aligned with the purpose, knowledge and skills developed through the course, and that are set out in the learning targets, assessment is likely to be valid and to exert a positive influence on teaching and learning.

DESIGN & TECHNOLOGY AND ASSESSMENT

Design & Technology (D&T) is a relatively new and developing subject. The purpose of learning D&T in schools is to enable students to be inventive in designing practical solutions to problems and so bring about change and improvements in existing situations. 'Designing' and 'making' are the two key skills to be developed in D&T studies. Through D&T activities, students are helped to acquire the capability to recognize and explore needs, to develop ideas about how these needs might be met, and to develop products that meet those needs. D&T capability assessment seeks to assess the extent to which students can use their understandings and skills when they are tackling a real and contextualised task.

United Kingdom situation

In recent years, the development of D&T led to dramatic changes in the content of the subject and the way in which it was taught. Assessment in D&T has been receiving more attention in the United Kingdom since 1979. Attempts were made by the Assessment Performance Unit (APU) to find ways to assess D&T capability (APU, 1987; Kimbell et al., 1991).

The APU posited a model of interaction between mind and hand as a new framework for assessment in D&T. In assessing the design skill, it emphasized the thinking and decision-making *'process'* that results in products rather than products that result from the processes. To assess this intellectual process, a number of strategies were developed to encourage students to make their 'intentions' explicit in action on a D&T task. APU split design activity and used short tests that covered a part of the process in the task. A holistic approach was adopted to identify students' capability in D&T rather than aggregating discrete levels of performance in areas like skills, knowledge and values.

Hong Kong situation

In Hong Kong secondary schools, some students choose to study D&T as one of their Certificate of Education (CE) or Advanced Supplementary (AS) level subjects at age 17 and 19 respectively. Traditionally, on completion of the study at CE or AS levels, public examinations are conducted to evaluate the learning outcome.

In D&T public examinations, both at the CE and AS levels, there are similar examination papers to test the different design performance of students. Paper One, 'design', in the examinations is a written paper. It is concerned with the ability to solve design problems with graphic solutions. The paper consists of one compulsory question and a few other short optional questions. Question number one is a compulsory item aimed at testing students' comprehensive performances in the process of designing. Other short items aim at testing limited design aspects or candidates' understanding of modern design.

REVIEW OF ASSESSMENT PRACTICES

In the past, assessment practices in Hong Kong have tended to be very quantitative and norm-referenced. Much criticism has been leveled against the use of norm-referenced assessment principles (Darling-Hammond & Wise, 1985; Koretz, 1991; Smith, 1991; Taylor, 1994). The current examination item design and assessment assumptions can hardly provide the necessary criterion-referenced information and the necessary room for a structured response to achieve the aims of learning D&T. Another mode of assessment is required to reflect the nature of achievement of students in D&T.

Quantitative assumption about assessment

In public examinations, the essential criterion to design assessment questions is based on psychometric considerations. Banding of students' abilities and fairness in marking of the test are the pre-requisite in developing test items. Therefore, the design problem was usually focused on confined design areas or confined technological aspects. In the examination paper, each sub-item of the design problem is allocated a certain proportion of marks. In marking the examination scripts, the marks will further be broken down into smaller units based on some criteria. On many occasions, the marks will be awarded quantitatively. For example, marks will be given on the number of alternatives of design proposals suggested. Marks will be assigned to each correct key design consideration spotted in the answers. Sometimes, marks will be awarded according to the quality of responses, e.g., to assign marks according to the feasibility of proposals or clarity of presentation. Finally, the scored marks in the sub-items will be added up as the total score. In the *Design Paper*, the comparison of design performance among students will be based on that total score. With the psychometric consideration as the primary concern, this traditional question format, marking scheme and scoring system in the current practice has limitations.

In D&T, especially at public examination level, students should be expected to be at a level where they could be given maximum freedom to develop their own ideas and to pursue any approach which seems to them to offer a reasonable outcome (DES, 1987; Siu, 1994). It is not appropriate to present the design requirements in the examination question in a prescribed pattern within very narrow limits. Students should be able to recognize problems to which a practical solution might be appropriate.

The current fragmented marking scheme and scoring system assumes that the answers to the sub-items require different discrete abilities. Certain proportions of marks were assigned for different design abilities, namely design analysis, design ideas, realization ideas, presentation skill, evaluation, etc. Assumption is made that the analysis or ideas generated under the prescribed considerations are equally important and significant. These discrete quanta of performances can be summed up to give an aggregate

score that yields an index of competence in design. In marking for correct answers against different categories, the marker gives a mark as each acceptable point is made, with ceiling bonus points for clever answers. This treatment of aggregating scores assumes the mutual equivalence, independence, and additivity among the different design abilities and the predetermined categories of responses.

From the design perspective, especially in an open-ended design situation, this assumption might not be appropriate to the assessment of total design performances. The format of question and scoring scheme under this assessment assumption limits the freedom of students to individual ideas, exploration and development. D&T emphasizes the learning outcome where students can build up the abilities to recognize problems, undertake inquiries by themselves, and contribute ideas accordingly. Different students can approach the same problem from different perspectives and produce different solutions. The mark allocation in the current examination items guarantees the homogeneity of students' responses, thus facilitating marking and comparison. However, students' efforts in attempting each portion of the question are guided by the mark allocation rather than the structured responses originated from students themselves. To a certain extent, this assessment mode defeats the aims of the D&T studies.

Assessment for norm-referencing

Besides the query of the quantitative assumption of the current question format and scoring, D&T assessment in the above situation is basically a norm-referenced test on the aggregated design abilities of students. In public examinations, the task is to discriminate between individual performances. A good test means the test can provide data for banding D&T students into A-grade, B-grade, C-grade and so on. It can be noticed that the test does not tell us how good an A-grade student is in relation to the D&T curriculum or predetermined standard. The mark scored or the band assigned in the test does not necessarily indicate a standard of attainment in design. The test item and the scoring system are not developed for this purpose. To D&T teachers, just to know the category of students' level in designing is not enough. It is far more important or useful to find out the extent to which students have learned the design skills and are able to apply them. D&T teachers should be able to determine whether the pre-determined standards levels have been met or not.

SOLO TAXONOMY – CRITERIA FRAMEWORK FOR JUDGING PERFORMANCE

Based on cognitive development theory, Biggs and Collis (1982) derived the SOLO Taxonomy. 'SOLO' stands for Structure of the Observed Learning Outcome. The outcome of learning can be revealed from the structured com-

plexity of performance. According to constructivism, learning grows cumulatively in stages in which the learned content is increasingly complex. In order to assess learning, to find out at what level a student is currently thinking with respect to a particular topic, it is necessary to be able to describe what the learning will be like at any particular stage.

Basically, there are two aspects to structural complexity: the amount of detail in the student's response (the quantitative aspect), and how well put together that detail is (the qualitative aspect). Both aspects are important, and can be classified by the SOLO Taxonomy (Biggs & Collis, 1982, 1989; Biggs, 1995).

Several levels define increasingly higher quality: increasing complexity of structure, abstractness, economy or elegance of processing, and originality of the response. In using the SOLO Taxonomy, there are five basic levels that might appear in statements to describe the test performance from incompetence to expertise.

1. Prestructural: The task is not attacked appropriately and the performance is incompetent.
2. Unistructural: One or a few aspects of the task are picked up and used.
3. Multistructural: Several aspects of the task are treated as if they were separate.
4. Relational: The quantitative aspects become integrated into a coherent whole; this level is what is normally meant by an adequate understanding of the topic.
5. Extended abstract: The previous integrated whole may be conceptualized at a higher level of abstraction and generalized to a new topic or area.

ALTERNATIVE ASSESSMENT

The SOLO Taxonomy has strong face validity that can be adopted in assessing the design performance of D&T students. Problems presented in D&T are usually ill-defined and complex, there is hardly a definite correct answer to be expected. In addition, the structural complexity of students' responses in the script can be assessed by target criteria of learning in D&T. There is also a need to establish a measurement method that encourages higher cognitive outcomes, and to be able to assess the design performance of D&T students.

An exploratory research was conducted to study question number one of D&T Paper One – 'design' at CE level. The study is based on an alternative approach by using the SOLO Taxonomy to assess the students' design performances.

Methodology

First, the focus was placed on the designing and scoring of the test items with different assumptions for different assessment approaches and their implication in studying D&T. Two test items and their respective marking schemes were used in this research. The homogeneous format of the 'Traditional Item' to restrict response and its quantitative nature in marking are considered contrary to the ultimate aims of D&T study. A 'SOLO Item' with maximized problem space was designed to enhance higher level outcomes of students. The design performance of students was assessed according to the level of structural observable outcomes displayed in the scripts.

Second, the focus was placed on assessing the performance of students in the tests. In order to study the difference of performance in attempting the traditional test item and the SOLO test item, subjects were required to complete both tests. The population of the study consisted of those who had D&T or related experiences at senior secondary levels or above. The subjects in this study included 79 students. One test item needed one hour and fifteen minutes to complete.

Lastly, the focus was to identify the measuring properties of the test items and marking schemes. Three experienced D&T teachers were invited to mark the scripts. A pre-marking meeting was conducted to discuss the marking schemes and to clarify any misunderstandings. Scoring sheets were provided for mark entry. The three markers worked independently and no writing on students' scripts was allowed. Each marker marked all the scripts of the whole sample, thus reducing the markers' variance in scoring. The adoption of multi-markers also enabled the correlation test of inter-marker variance to be conducted. Scores from the two test items and their respective scoring schemes were then collected for analysis (Table I). Subsequently, follow-up interviews were conducted to obtain feedback from students to clarify some of the uncertainties in this study.

TABLE I
Scores from different test items and scoring schemes

	Traditional Item	'SOLO Item'
Traditional scoring	* TT score	ST score
SOLO scoring	TS score	** SS score

Note: * Scores marked by two markers (researcher and one additional marker).
      ** Scores marked by two markers (twice by researcher and one additional marker).
      'TT' Traditional item marked by traditional scheme.
      'TS' Traditional item marked by SOLO scheme.
      'ST' 'SOLO Item' marked by traditional scheme.
      'SS' 'SOLO Item' marked by SOLO scheme.

The content to be assessed

The *APU Framework of Assessment of Performance in D&T* was adopted and modified in this study. It included the 'process' and 'holistic' approach being used in the assessment.

The *design skills* used in design and technology activities are the skills constituted in the process of recognizing a need and matching available means with desired ends. They can be grouped into five categories:

- Investigation – abilities to identify 'needs' and 'opportunities' for designing;
- Specification – abilities to determine design statements for designing;
- Solution generation – abilities to generate ideas to match design requirements;
- Planning for making – abilities to plan for realizing the design ideas in terms of production activities, resources application and quality specification;
- Evaluation – abilities to make judgments on design.

Technological knowledge concerns the ability that knowledge and understanding on control, energy and materials can be applied in specific practical problems to the creation of a device or a system which works. This ability was reflected in the students' response in solving the design problem. Marks were given for the correct application of technological knowledge. Higher SOLO levels were awarded when more pieces of technological knowledge were observed and they were related together correctly and wisely to solving a problem.

Value judgments is the ability that appropriate reasoning on technical, economic, aesthetic and moral aspects can be exercised in the activities related to design and technology. This ability was also reflected in the students' response in solving the design problem. Marks were given for correct judgment and decision made. Higher SOLO levels were awarded when more judgments and decisions were observed and they were reasonable and exhibited the well-organized solving of a problem.

Traditional Item – prescribed response

The current format and marking scheme in the D&T public examination are considered traditional. An examination question in the past paper of 1994 was adopted in this study. The item asked for the designing of a set of folding chairs for a bus station. The solution to provide a comfortable facility for the waiting passengers was predetermined in the given brief as a folding chair but not others. The five prescribed areas of design considerations were listed and candidates were required to further analyze and make design proposals accordingly. The question made explicit demands on the candidates to show the detail of the method of fixing the chair seat to the bus shelter.

'SOLO Item'– open-ended response

The newly developed item should be able to provide maximum freedom for students to develop their answers and it was designed to be marked by the SOLO Taxonomy. Based on the new assumptions, a new 'SOLO Item' was designed. The item asked students to design a device or system to collect tennis balls that were spread over the ground. The brief provided contained no prescribed design requirement or response format except requiring a design solution in terms of a device or system. Although students needed to use both specific technological knowledge and design values in tackling the problem, neither one of them was mentioned nor demanded in the question. Students were expected to define and use suitable knowledge and values whenever they were justified. Instead of asking specific tasks in the sub-items to guide the desired outcomes as in the traditional item, the 'SOLO Item' reminded students that assessment would be made holistically on the various design skills displayed in their answers. The design skills were mentioned in a broad sense, in line with the question expectations described in the examination syllabus. Students were left free to determine the way that the question could be approached themselves.

Traditional marking scheme

A marking scheme similar to that used by the public examination was adopted in this study. In addition to those maximum marks allocated in the sub-items of the question, the marks were further broken down into smaller detailed units for specific responses. The maximum allocation of marks to each area represented the relative importance of the performance to be assessed. The aggregated score of all the small units became the indicator of total design performance for comparison.

SOLO marking scheme

Biggs' system of five-level SOLO Taxonomy was adopted in this study. As a general guideline, the five levels of SOLO Taxonomy on D&T design performance are represented in Table II.

FINDINGS

This study provided both quantitative and qualitative evidence of the test designs and the test effects on students' design performance in D&T.

Design performance of students

A summary of assessment result according to SOLO Taxonomy in the 'SOLO Item' is described in Table III.

Considering the overall design performance of students, a full range of

TABLE II
SOLO level description

SOLO level	Description of performance
1. Pre-structural	Display incompetence in design. Problem is not attempted or the key aspects are not clearly defined or solved. Fail to relate the ideas presented to the problem. Information produced does not benefit solving the problem.
2. Uni-structural	Display limited design abilities. Problem is defined from a narrow perspective at a superficial level. One or a few aspects are picked up in designing. Some important aspects are missing in the design ideas. Although not many aspects of exploration and judgment are observed, they can lead to weak or simple solutions to solve problem with minimum quality.
3. Multi-structural	Display comprehensive design abilities, but in isolation. Problem is defined from wide perspectives with many design ideas generated. Essential and important aspects are picked up in designing. Many elements of exploration and judgment are observed. However, the design ideas are loosely organized, with different ideas not integrated coherently. Some design features misfit another, and judgments are not consistent.
4. Relational	Able to relate different design skills to form coherent analysis, statements, design ideas and judgments. Answers are not only a sound design proposal to the problem, they are presented in a coherent and structured way. Explanation of why and how the solution is developed, realized in practical terms, and evaluation judgments on how far the solutions satisfy the original needs and specifications, are components.
5. Extended abstract	Display higher modes of operation in structuring knowledge to solve a problem. In addition to what can be observed at the relational level, some new and creative ideas through logical and mature design developments are presented.

TABLE III
Results of the 'SOLO Item' assessed by SOLO taxonomy

	Design profile					Overall Performance
	Investigation	Specification	Solution	Planning	Evaluation	
Prestructural	39 (50%)	43 (54%)	18 (23%)	49 (62%)	51 (65%)	**22 (28%)**
Unistructural	15 (19%)	10 (13%)	24 (30%)	17 (21%)	17 (21%)	**35 (44%)**
Multistructural	12 (15%)	15 (19%)	22 (28%)	11 (14%)	8 (10%)	**14 (18%)**
Relational	12 (15%)	10 (13%)	14 (18%)	2 (3%)	3 (4%)	**7 (9%)**
Extended – abstract	1 (1%)	1 (1%)	1 (1%)	–	–	**1 (1%)**
Alpha			0.68			–

Note: Sample size – 79 students.

SOLO levels from prestructural to extended abstract were observed in the students' scripts to solve the D&T design problem. The largest proportion of 35 students (44%) is at the unistructural level.

Although it was the first time most D&T students attempted the 'SOLO Item' format, no specific problem was observed in the test administration. Some students with no experience in both test item formats were found to have high performances in the newly designed 'SOLO Item'.

The internal consistency of design profile among various aspects is at an acceptable level. The coefficient alpha among the profile scales is 0.68. It was considered that the new item format and the SOLO-scoring scheme had potential as an alternative approach to assessing the design performance of D&T students.

Correlation among the assessment dimensions

Correlation analysis was used to study the relationship between and across the nature of item designs and the scoring schemes. In this analysis, the two test items are assumed to be a pair of parallel tests. The item context is similar: both are performance tasks, paper tests, and practical problems. A comparison between the two items according to the item natures and assessment models is shown in Table IV. The result of correlation coefficients among different scorings is shown in Table V.

TABLE IV
Comparison of test items according to dimensions and modes of assessment

Item context	Situated task – Tennis ball collector		Situated task – Folding chair set	
Item nature	SOLO/Quantitative and Qualitative (open-ended to encourage max. freedom of context exploration and response organization)		Traditional/Quantitative (prescribed response in context and organization)	
Assessment model	SOLO/Criteria-referenced on thinking	Traditional/ Quantitative scoring	SOLO/Criteria-referenced on thinking	Traditional/ Quantitative scoring
Score	SS score	ST score	TS score	TT score

TABLE V
Result of correlation coefficients among different scores

	TS	TT	ST	SS
TS	1.00			
TT	0.66	1.00		
ST	0.34	0.28	1.00	
SS	0.28	0.14	0.56	1.00

If the correlation is close to '1', the two assessment models may measure the same ability. If the correlation is close to '0', the two assessment models may measure completely different abilities. Testing by the same item (same context and nature), the correlation between the two abilities is quite steady at 0.56 ('SOLO item': SS to ST) and 0.66 (Traditional item: TS to TT). The SOLO score and the traditional score on the same test represent two related, but different, abilities of students. The two marking schemes are measuring two different abilities. It supports the assumption that the SOLO Taxonomy is sensitive to the structural complexity of observable outcomes, and the traditional scoring is sensitive to quantitative responses.

The relatively low correlations of the SOLO Taxonomy and traditional marking scheme in the two different items are 0.28 (ST to TT) and 0.28 (SS to TS) respectively. Almost no relationship (0.14, SS to TT) was found between the different assessment models on items with a different item context. These correlation analysis indicated that SOLO scoring and traditional scoring are measuring similar but different abilities in a D&T design task. The more the assessment dimensions varied, the less the relationship was found among the scores.

The qualitative data collected in the interview with students provided further evidence of the above analysis. Students agreed that the SOLO scoring assessed more the thinking abilities in tackling design tasks while the tradition marking scheme measured more knowledge-based designing. Most students perceived that the traditional item was more factual, demanding more technological knowledge. On the other hand, the 'SOLO Item' was perceived as requiring more thinking and organization skill.

Tests give information on what is needed to handle the test. Backwash refers to the fact that testing drives not only the curriculum, but teaching methods and students' approaches to learning (Biggs, 1995). If the new approach in this study were adopted in the daily teaching and examination situations, the backwash effect of assessment could possibly help students towards higher cognitive outcomes in design performance.

Reliability of marking

Two studies were conducted on the reliability of marking using the two marking schemes. One compared the inter-marker variance and the other checked the consistency in marking. Correlation coefficient was used as a reliability indicator of these studies. The results are shown in Table VI.

A relatively low correlation of 0.49 between different markers in the SS score was observed. It indicated that the reliability of using the SOLO Taxonomy as the marking scheme between markers was low. However, a higher mark-remark consistence of 0.71 by single marker in the SS score was recorded. This correlation coefficient was the same as the inter-markers' variance in the TT score.

As a first attempt to use the SOLO Taxonomy, this level is not satis-

TABLE VI
Correlation coefficient between markers

Score	Markers	Correlation
SS	2 markers	0.49
	same marker	0.71
TT	2 markers	0.71

factory, although acceptable. From the perspective that assessment assumptions are grounded on teaching and learning, reliable measures are not the only consideration for measurement. A valid test is equally important and even more critical in assisting teaching and learning.

In order to implement this new assessment approach to the fullest extent, preparing teachers to adopt the testing mode, and training the skills in making decision for assessors are deemed to be essential. If more effort and resources were allowed, the reliability of the tests could be improved considerably. For example, with suitable moderation measures and greater experience and improvement in developing precise grade descriptions with illustrated examples, it is believed that there is room to increase the reliability of marking using the SOLO framework.

Validity of test design

Very positive and encouraging feedback on the new item design and the SOLO level criteria in design performance was received from students. Students supported the new item format to allow more freedom, opportunities and design space to answer. Most students agreed that the feedback of their SOLO level truly reflected their thinking performance in the test. A majority of them preferred to have feedback from the SOLO level to help them learn D&T. They considered it useful that the SOLO levels could help them formulate their learning goals.

Construct validity of the new assessment approach was therefore established in this study. Students generally perceived the assessment assumptions of the newly-design item format and marking scheme as demanding more structured thinking and higher analytical power. Some capable students in fact had a strong preference for this item format. When the SOLO criteria on design performances were explained and the level achieved by students was reported, most students agreed that this assessment feedback on the design profile could be helpful for their learning. The targeted level helped them to formulate their learning goals.

To obtain further evidence of validity, ongoing investigations of the consequences of the interpretation and use of this alternative assessment approach must be conducted (Messick, 1989). In addition, background variables that affect performance must be investigated. Evidence was determined as being required to validate the backwash effect of the assessment if the

item were implemented in the public examination situation. The treatment effects when teachers used a different strategy to tackle the new assessment design needs to be taken into consideration.

CONCLUSION

This exploratory study attempted and demonstrated, not only the feasibility of using the SOLO Taxonomy for assessing D&T work, but also the reliability and validity of such an assessment technique. Driven by the higher level of thinking and creativity skills demanded of the subject, the open-ended SOLO type of design problem was suggested to be a more appropriate question format to elicit higher cognitive outcomes of students. The SOLO Taxonomy was suggested to be a more accurate indicator of students' design performance, as opposed to traditional marking. The feedback of design achievement to student was suggested to be a profile with graded levels using explicit criteria rather than just a banding grade.

REFERENCES

APU: 1987, *Design and Technological Activity: A Framework for Assessment,* Assessment of Performance Unit, DES.
Biggs, J. B.: 1995, 'Assumptions Underlying New Approaches to Educational Assessment: Implication for Hong Kong', *Curriculum Forum* **4**, 1–22.
Biggs, J. B. & Collis, K. F.: 1982, *Evaluating the Quality of Learning: The SOLO Taxonomy,* Academic Press, New York.
Biggs, J. B. & Collis, K. F.: 1989, 'Towards a Model of School-Based Curriculum Development and Assessment: Using the SOLO Taxonomy', *Australian Journal of Education* **33**, 149–161.
Darling-Hammond, L. & Wise, A. E.: 1985, 'Beyond Standardization: State Standards and School Improvement', *Elementary School Journal* **85**, 315–336.
DES: 1987, *Craft, Design and Technology from 5 to 16,* Curriculum Matters 9, HMI.
Kimbell, R. A., Stables, K., Wheeler, T., Wozniak, A. & Kelly, V.: 1991, *The Assessment of Performance in Design and Technology,* APU project. Evaluation and Monitoring Unit. School Examination and Assessment Council (SEAC), London.
Koretz, D. M.: 1991, *The Effects of High Stakes Testing on Achievement: Preliminary Findings about Generalization Across Tests.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.
Messick, S.: 1989, 'Validity', in R.L. Linn (ed.), *Educational Measurement*, Macmillan, New York, 13–104.
Siu, K. W. M.: 1994, *A Study of Pupils' Rationale or the Selection of Topics in the Project Selection of the HKCEE Design and Technology,* University of Hong Kong, M.Ed. Dissertation.
Smith, M. L.: 1991, 'Meanings of Test Preparation', *American Educational Research Journal* **28**(3), 521–542.
Taylor, C.: 1994, 'Assessment for Measurement or Standards: The Peril and Promise of Large Scale Assessment Reform', *America Educational Research Journal* **31**, 231–262.