ASCD

REVIEW

Vol 3 No 1 Nov 1992 MITA(P) NO 572/09/92

Assessment

Association for Supervision and Curriculum Development (Singapore)

ASSOCIATION FOR SUPERVISION AND CURRICULUM DEVELOPMENT

EXECUTIVE COUNCIL

PUBLICATIONS COMMITTEE

ASCD (Singapore) Review is published three times a year in March, July and November. The views expressed in this journal do not necessarily reflect the official position of ASCD (Singapore).

The Publications Committee seeks articles and letters that provide useful information on the the teaching/learning process. Manuscripts should show the author's name, title and institution. Contributions should be in the form of a hardcopy together with a 5¼ or 3½" diskette. Please send all contributions to the Publications Committee, ASCD Singapore, c/o CDIS, 465E Bukit Timah Road, Singapore 1025.

Published by Association for Supervision and Curriculum Development (Singapore).

Printed by Mentor Printers Pte Ltd, Blk 155, Kallang Way #01-03, Singapore 1334.

Association for Supervision & Curriculum Development (Singapore)
REVIEW
November 1992 Vol 3 No 1

FOCUS: Assessment

OTHER TOPICS

BOOK REVIEW

LOW GUAT TIN

A Chat with Professor Madeline Hunter

First we are told by Rousseau in his book Emile that "To lose time is to gain time" and that was in the 16th century. Today we are told in another book that "To teach less is to teach more." Like Rousseau's statement "Less is more" sounds conflicting. How can one teach less yet teach more? How can one lose time to gain time? The master teacher, Professor Hunter explained it thus:

If a student knows something really well, it will transfer more accurately and more predictably into a new situation. Therefore it is better to teach one thing really well so that the student can use it all his life than it is to cover two or three things that the student doesn't understand and as a result that information will not transfer.

But what about the syllabus? We have a syllabus to cover and we are accountable for that, argued the interviewer. Covering the syllabus is a serious business to many teachers and many are ever so concerned that they have not covered all that is in the syllabus. The wise lady looked at the persistent Singaporean and said:

But if a student doesn't learn it, what good has covering the syllabus done? We act as if when the syllabus is covered, everybody knows everything in it. Yet a lot of students don't. We can find that out in reading or vocabulary. If you teach 5 words that a student can learn in a meaning loaded sentence and use about himself, that student will use those words for the rest of his life. On the other hand, if you make a student memorize the definitions of 20 words, then after the student passes the test he will forget the words and they will never be used again. That's why *teaching less is teaching more.*

The interviewer persisted. "Yes, but in Singapore we have examinations - major examinations - and this is an exam-oriented society. If you teach less, wouldn't you be short-changing students? Anyway, would you say that examinations force us to narrow the curriculum?" This time the Professor replied:

No way. The thing about exams is that we are finding out what students don't know. So when we teach something really well, that's going to be remembered longer and transferred more accurately to new situations rather than just running through it, cramming for a test - and forgetting it. The whole purpose of assessments or examinations is to help a student recognise what he knows and what he doesn't.

Not one to keep quiet, interviewer interrupted. "But we have external examinations. A child takes about 4 external examinations from Primary 1 to JC 2. External exams are big issues which every school has to grapple with. What are your views about such examinations?" The master teacher smiled and said:

There's nothing wrong about that if they are used for the purposes intended. Your "O" level examination spreads students out. It tells you who are your best students, who are your weaker ones and that's something you may need to know. But it doesn't help a teacher know what she should be teaching that student tomorrow morning. It's perfectly okay to have

The whole purpose of assessments or examinations is to help a student recognise what he knows and what he doesn't.

We strongly recommend that in his teaching, a teacher does dip sticking - seeing that there is enough knowledge before going on and that means that teachers must have a good idea of students also.

external examinations. In the States we have entrance examinations. We even have examinations for doctoral students.

The interviewer persisted, "Many teachers worry about the final examinations every year, they want their students to do well. How can we ensure that students are learning, i.e., they are not merely cramming and after the exams everything is forgotten?" She lighted up and said:

The way to do that is by increasing skills in teaching - the only way to do that. Doesn't matter how you organise, how you test etc. It depends on one factor - how skilled that teacher is in teaching. That is the critical element and we have done everything else to change the skill of the teacher. In fact within the last 20 years, we have been looking at the process of teaching and saying that if a teacher does this or that, kids can learn more. For example, if a teacher was signalling during teaching, then the teacher stops and clears up the confusion right there and then. No kid can then pretend they know and become more confused. The teacher clears it up there and then through signalling or through a written test.

(Note: signalling - students make signs to the teacher to indicate whether they understand, e.g., thumbs up to show that they understand. Or the teacher could give a problem sum and students indicate whether the process involves division or subtraction. Signs or signals are previously agreed upon. The aim is for students to send periodic messages to teachers about how well they are understanding.)

But we have over tested children. This is why we strongly recommend that all the way in his teaching, a teacher does dip sticking - seeing that there is enough knowledge before going on and that means that teachers must have a good idea of students also.

(Note: Dip Sticking is to ask questions around the class, i.e., to pick on pupils all over and asking lots of questions to check for understanding. The aim is to monitor continuously whether students understand.)

The interviewer persisted. Another issue we have in Singapore is we use examination scores to rank schools. Schools are ranked based on their "O" level scores. What are your views on this? The master teacher replied instantly:

We do that for university entrance examinations too. Our scholastic aptitude tests (SAT) do the same kind of thing. That's okay if you want to rank people and spread them out. If you say school A outperforms school B, does it tell you that the teaching in school A is better? Students from school A could be from professional families, affluent families and students from school B from manual labour homes and it doesn't tell you how good a job teachers are doing. School B teachers might be doing a better job at teaching. Often times very bright children learn in spite of what you are doing. In a way it's like saying if an affluent child gets a really good breakfast and dinner, if he has a poor lunch it doesn't make any difference. The poverty child who has a poor breakfast and a very poor dinner, if he gets a very good lunch, that will help him close that gap.

"Still on the subject of assessment," the interviewer droned on, "in your talks you mentioned authentic assessments. Can you tell us a bit more?"

We are really assessing what a student has learned, what he is ready to learn next rather than just testing him on much lower level things. We are finding out that standardized tests measure very accurately what they were designed to measure but they were designed to measure where the child fits with the group that's been normed, in other words a comparison group. If a student meets the comparison group, that's fine. However standardized tests were used for things they were never intended. It's

fine to compare programmes of school district A with B and so forth but as far as knowing what to teach a particular student tomorrow morning, they are not good for that. A criterion referenced test often replaces them because in a criterion referenced test, it tells you what a student can do with a particular objective or chore, e.g., can he write a persuasive essay?

Authentic assessment is a different thing. It is a timed thing where presence or absence of success doesn't depend on the fatal moment, e.g., we would have a student write a persuasive essay that gives a baseline, say to persuade his teachers not to give him so much homework.

Then you would teach persuasive essays where the criterion measures are:

- Can a student present a point of view?
- Can a student support it with objective evidence, not just opinions?
- Did the student anticipate a counter point of view?
- Was it all presented in a cogent argument with an introduction and a concluding paragraph?

In other words we can differentiate between the child who says "I don't think we should have so much homework, I don't like it" to one who presents a persuasive argument about other kinds of important things in life that he had to give up and so forth.

Authentic assessments contain actual products of the child's work. There are 3 ways of knowing:
1 Can you generate the knowledge, skills, product, procedure whatever it is?
2 Can you recognise a correct one when you see it?
3 How long does it take for relearning?

So we are not only measuring recognition which standard achievement test has been measuring but the ability to generate views yourself. The advantage is that it gives you measurements over time from the beginning to the middle to the end of learning. Give us check points in between rather than deciding on the fate of students in one day where he may have been sick or where he does not produce his typical performance.

The interviewer went on, "How does this authentic assessment work? That is, do teachers grade all the work, is it like continuous assessments?" The professor was quick to reply:

Yes, but you select what you put in the portfolio. The teacher and the student select what to put in. One student's portfolio could be different from that of another. One who is a visual learner will have diagrams etc and the other may have an essay.

What is a porfolio?

A portfolio is a collection of student's generated work, e.g., an essay can be in a portfolio, a page of maths and so on. A portfolio is basically to help the student and his teachers assess learning. A student has to ask himself "What have I learnt? What am I ready to learn?" This is more effective than for a student to say I'm an A or a B student. Because if we say we are an A student and we are good but good in what? Portfolios say I'm very good in this, here's something I need to improve. Yes, this is an area that I'm not so good in, but by and large I'm a good student.

In fact we have students very involved in selecting materials for their portfolios and often times determing what it is to be measured. What would be a fair measure of their learning? The students become very involved in their learning rather than becoming passive recipients.

Puzzled, the interviewer asked, "Does this form of testing assess student's academic achievement

A criterion referenced test tells you what a student can do with a particular objective or chore, e.g., can he write a persuasive essay?

reliably?"

Depends on what is selected. Like every other test it depends on how it is done. You can get the best achievement test in the world and you have people teaching to the test...anything is only as valid as the way it was done.

We are still not satisfied. Portfolios are now still randomly assembled. We are just learning what are the essentials in portfolios and what aren't. It's an important learning, for it's causing teachers to say what evidence would there be that this child can now do well, something that he couldn't do before. There is now objective evidence. The problem of portfolio is the selection of evidence. What selection of evidence to show that the person is wanting of one thing or another.

How does one establish standards in portfolio assessments then?

We are in the process if establishing standards. We don't have them. Right now standards are being developed by people but we don't have a blanket standard yet. Portfolio is a very, very good idea but remember when they first built the U2, the first one crashed. We don't give up. We are working on it.

How do you ensure objective assessment in portfolios?

Create a protocol such as: Is the student expressing a point of view in support of the data? Is the student anticipating counter arguments and either deluding them or eliminating them. We establish a protocol for it. Is the student using a variety of sentence pattern? Is he using descriptive words and such likes.

If we were to use portfolios in Singapaore, how then do we compare schools?

If you create standard portfolios, it's just as comparable as anything else. At the moment we are comparing schools where the teachers are teaching very differently and we seem to be satisfied about that. We are also comparing schools where the bulk of the teachers are first year teachers with another where the teachers are very experienced and that doesn't seem to bother us.

Remember, portfolios are not the be all and end all. We are not eliminating other kinds of testing. We are now teaching for transfer over and beyond just understanding. Can students use it in new situations? This is the essence of thinking. We want kids to think. They can't think unless they transfer. We want kids to generate responses.

Because there is such a variety of presentation, how does the teacher grade them?

The issue is "Does it demonstrate understanding?" Can you grade a poem? Can you grade a sonnet? Now poems, sonnets are all different but the teacher can grade them. Does it have critical attributes of a sonnet etc. In a doctorate, people don't do all the same dissertation but we can still grade them. We can pass them or pass them with Honours.

The idea that everything has to be identical doesn't hold. For instance, we both go for a medical check-up and we have different blood pressure, different weights and so forth but we are both certified healthy. The same doctor judged that we are both healthy but we don't have identical health records.

Yes, indeed the issue in this day and age is understanding and transfer. Students may score in PSLE, O or A level exams, but can they transfer? With that I hope we understand why covering the syllabus may not be all that crucial for often in our bid to cover the syllabus, we actually "covered" it for many of our weaker students. There is thus a lot of wise adage in these words "Less is More."

Portfolios are not the be all and end all. We are not eliminating other kinds of testing.

Madeline Hunter is Adjunct Professor of Education, University of California at Los Angeles, USA.

Low Guat Tin is Senior Lecturer, Policy and Management Studies Division, National Institute of Education, Nanyang Technological University, Singapore.

SOH KAY CHENG AND POH SUI HOI

Assessment: Uses and Training Needs

A medical doctor who prescribes but does not diagnose is unimaginable, because diagnosis is the basis of prescription. Here, diagnosis and prescription are considered as two stages of the same process of improving the patient's physical and mental well-being. In teaching, much emphasis is placed on teaching *per se* rather than on a balance between teaching and assessment, and assessment is not infrequently seen as a necessary evil that takes away the joy of the teacher's teaching and the student's learning. This is not the fault of anyone involved in teaching as the semantic of the word **teaching** strictly means imparting knowledge, training skills, and moulding character in the learner; assessment is not part of the meaning. Conventionally, teacher education programmes do not pay sufficient attention to developing teacher-trainees' understanding and skills relevant to assessment; the tacit assumption is that if they have been trained to teach, the novice teachers will learn and assessment is only an adjunct to the training process. However, in recent years it has been recognized that assessment is and should be seen as part and parcel of teaching and hence training assessment skills should form an integral part of teacher education. The close relationship between teaching and assessment is succinctly portrayed by the following quotation from a recent article, *High Quality Classroom Assessment: What Does It Really Mean?*

> The quality of instruction is a function of teachers' understanding of the strengths and weaknesses of their students. The depth of that understanding, in turn, depends on the quality of teachers' assessment of student achievement. Thus, sound instruction requires sound classroom-level assessment of student achievement. (Stiggins, 1992: 35. Emphasis added)

Earlier, Schafer (1991) identified two reasons for deficiencies in measurement education for many teachers in the United States: (a) ineffective communication of the importance of assessment concepts and methods for effective teaching and (b) lack of clear expression on the part of the measurement community about what should be included in that training. He went on to suggest the following eight content areas of essential assessment skills in the professional education of teachers:

- Basic concepts and terminology of assessment
- Uses of assessment
- Assessment planning and development
- Interpretation of assessment
- Description of assessment results
- Evaluation and improvement of assessment
- Feedback and grading
- Ethics of assessment

The concern for assessment in the American scene in recent years is further reflected in a large scale study of uses and abuses of achievement test scores (Nolen, Haladyna & Haas, 1992). The survey involved more than two thousand teachers and administrators at the primary and secondary levels. Secondary teachers perceived the school administrators as using test scores mainly for advertising the schools, evaluating school effectiveness, and in "highest score" competition; but teachers themselves used test scores mainly for identifying remedial students, measuring class/school effectiveness, identifying gifted students, placing students for instruction, communicating with

Conventionally, teacher education programmes do not pay sufficient attention to developing teacher-trainees' understanding and skills relevant to assessment

In Singapore, the results of assessment have long-range impacts on individual student's academic advancement as they are used for placement throughout the whole educational ladder.

parents, and diagnosing learning problems. Obviously, the same information obtained through assessment have been used for different purposes by different school personnel with different interests.

The critical role of assessment in our education system here in Singapore cannot be over-emphasized for several reasons. As is well-known, results of assessment have long-range impacts on individual student's academic advancement as they are used for placement throughout the whole educational ladder. Secondly, the authorities and the public judge schools (though not solely), by their performance on national examinations, as evidenced by the recent release of the **ST Schools 100**. Thirdly, it is a common practice that assessment is used as a mechanism to motivate students to learn. Last but not least, much of the teachers' time is used for professional activities related to assessment - setting test papers, marking students' answers, collating marks, and reporting results. It is therefore useful to find out in what ways test results have been utilized for various educational purposes by both the school administrators (i.e., principals, vice-principals, and HODs) and the classroom teachers and, in relation to this, what training needs are required by teachers for a more efficient discharge of this part of their responsibility.

Table 1
Administrator Use of Test Scores and Exam Results

Uses	%
Identify curriculum strengths/weaknesses	79.0*
Evaluate school effectiveness	72.8
Evaluate teacher effectiveness	52.5
Advertise the school	52.5
Evaluate teaching methods	50.6
Identify teacher strengths/weaknesses	43.8
Evaluate materials (e.g., textbooks)	22.5

* Routinely and Often

An opportunity for such a study was afforded by the participation of about 100 experienced teachers from more than seventy secondary schools in the Further Professional Diploma in Education Programme (FPDE-S) at the National Institute of Education. The programme prepares the participants for the role of a Head of Department (HOD) in several subject areas. As HODs are expected to provide leadership in assessment, the module *Classroom-Based Evaluation* was specially designed to meet this need by training the prospective HODs in principles and skills relevant to assessment of student learning. The content areas of this module are as follows:

- Purposes of evaluation in the classroom context
- Instructional objectives and evaluation procedures
- Evaluating cognitive learning: selection-type items
- Evaluating cognitive learning: supply-type questions
- Evaluating affective learning: attitudinal scales
- Evaluating psychomotor learning: rating and observation
- Improving tests through item-analysis
- Describing test performance
- Ensuring test quality
- Reporting test performance

Based on the content covered in the *Classroom-Based Assessment*, a questionnaire was designed with reference to the professional training needs in assessment identified by Schafer, 1991; Stiggins, 1992 and the survey by Nolen, Haladyna & Haas (1992). The questionnaire was administered at the end of the course to the prospective HODs, numbering

Table 2
Teacher Use of Test Scores and Exam Results

Uses	%
Identify remedial students	97.5*
Guide instruction	82.7
Place students for instruction	79.0
Diagnose learning problems	75.3
Communicate with parents	75.3
Predict student performance	74.1
Measure class effectiveness	70.0
Evaluating teaching methods	54.3
Stimulating curriculum review	51.9
Identify students for special services	34.6
Identify gifted students	31.6

* Routinely and Often

Table 3
Teachers' Training Needs in Assessment

Areas of training	Importance	Urgency
Planning assessment	53.1*	42.0*
Interpreting assessment results	34.6	38.3
Evaluating test quality	33.3	48.1
Developing essay-type questions	32.1	29.6
Developing objective tests	29.6	32.1
Uses and limitations of assessment	27.2	18.5
Concepts and terminology of assessment	23.5	19.8
Communicating assessment results	23.5	14.8
Grading essay-type answers	19.8	18.5
Using item-analysis techniques	17.3	19.8
Understanding effects of testing	17.3	19.8
Developing rating scales	6.2	9.9

* First three ranks.

near one hundred and coming from more than seventy secondary schools. Eighty-one usable responses were used for this analysis. The prospective HODs come from several disciplines covering a wide range of subject specializations: Language & Literature (22.5%), Science (20.0%), Mathematics (18.8%), Social Studies (5.05), Technical (8.8%), ECA (12.5%), Chinese Language (6.3%), and Media (6.3%).

Table 1 shows a large majority of respondents reporting that school administrators routinely or often use test scores and exam results for *identifying curriculum strengths and weaknesses* and *evaluating school effectiveness*. About half reported that school administrators use test and exam information for *evaluating teacher effectiveness, evaluating teaching methods*, and *advertising the school*. A slightly lower proportion reported using the information for *identifying teacher strengths and weaknesses*. A much smaller proportion reported its use for *evaluating materials*. It appears that the school administrators have been using information from assessment for the broader issues of curriculum suitability and school effectiveness more than for evaluating individual teachers and specific instructional materials such as textbooks.

As classroom teachers have a different perspective from that of the school administrators, the same assessment information obtained through tests and examinations would have different functions. As shown in Table 2, the majority of the prospective HODs reported that teachers have been using assessment results for *identifying remedial students, guiding instruction, placing students for instruction, diagnosing learning problems, communicating with parents, predicting student performance,* and *measuring class effectiveness*. About half of the HODs reported teachers using assessment information for *evaluating teaching methods* and *stimulating curriculum review*. A smaller proportion of HODs reported teachers' use of as-

sessment for *identifying students for special services* and *identifying gifted students*. As gathered from the responses, it is obvious that teachers' use of assessment has its focus mainly on the students and their learning difficulties. It plays a less prominent role in review of teaching methods and curriculum and much less in identifying students for special educational needs.

Of the twelve areas of training needs (Table 3), *planning assessment* heads the list in importance, followed closely by *interpreting assessment results, evaluating test quality, developing essay-type questions,* and *developing objective tests*. It is to be noted that four of these five areas identified also appear as 'high' in training urgency. Training in the more conceptual aspects of assessment were considered less important and less urgent. Similarly, technical training in grading essay-type answers and item-analysis of objective items were considered 'low' in importance and urgency. It seems that the prospective HODs perceive that the classroom teachers' more immediate training needs lie with planning, constructing, and using assessment. This suggests that in the minds of the prospective HODS, conceptual understanding and technical skills can wait.

Over the past few years, campus-based in-service courses on testing and measurement have been conducted for teachers who wish to improve their theoretical knowledge of and practical skills in assessment. It was one of the more popular in-service courses, although its impact has yet to be evaluated. As training needs may be met through various arrangements and probably with different impact, a question was asked about the most effective arrangement for assessment training for teachers. Table 4 shows that the prospective HODs perceived school-based workshops as the most effective training arrangement, conducted either by NIE staff or by the joint effort of NIE staff and HOD in the school. Campus-based in-service courses follow next, with about a third of the HODs attesting to this method. The responses also show that research-related activities were seen as least effective in meeting teachers' assessment training needs.

Table 4
NIE Lecturers Meeting the Teachers' Training Needs

Training arrangements	%
Conduct school-based workshops	65.8*
Joint effort with HOD in school-based workshops	48.1
Conduct campus-based inservice courses	32.9
Serve as resource persons for consultation	21.5
Assist in school-based research on assessment	17.7
Research on assessment and share the findings with the schools	13.9

* First two ranks.

The present study has not uncovered anything unusual; it merely presents what seems to be common knowledge in a concrete form. The fact that the same assessment information is used for different purposes by incumbents of different positions in the schools is not surprising and is similar to that found in the United States. It helps to prioritize the needs of teachers in assessment training and the modes of training seen as likely to be more effective. Too often, the training programmes planned mismatched the needs of teachers (Schafer, 1991; Nolen et al., 1992). This certainly would result in a lot of wastage in terms of time, money and effort. As suggested by the findings, teachers' assessment capabilities can be developed through school-based workshops conducted by NIE staff, either by themselves or jointly with HODs. Campus-based in-service courses could also be looked into. The content areas in the training programmes should cover first, the more practical and technical aspects of classroom assessment, leaving the more conceptual and theoretical aspects to later occasions.

Research suggests that assessment-related activities take up as much as one-third to one-half of the available

School administrators routinely or often use test scores and exam results for identifying curriculum strengths and weaknesses

Research suggests that assessment-related activities take up as much as one-third to one-half of the available professional time of American teachers - the situation here is not that different.

professional time of American teachers (Stiggins, 1992). The situation here is not that different - our teachers could very well be doing more than their counterparts elsewhere. In view of the critical role assessment plays in our education system and the substantial amount of time consumed for assessment-related activities, there is obviously a need to enhance the teachers' capabilities in assessment and to find time-saving devices for more productive use of teachers' available professional time. There are, however, other aspects closely related to assessment that schools may explore. Here are some possibilities worthy of consideration within the school context.

It is well-known that setting and marking test papers are labour intensive, time-consuming, and repetitive. Perhaps, all three are closely related. Therefore, re-cycling items by item pooling either within a school or across a few schools in a certain zone, is not only time-saving but has other benefits as well. This idea has been around for quite some time and some schools are actually doing it. What can be added is the use of personal computers (PCs) for (1) marking objective questions, (2) doing item-analysis, (3) storing and retrieving analyzed items. The value added to computerized marking and item-analysis is obvious. The value added to the re-cycling of items is that it enables comparisons to be made, where desired and desirable, between years within-school and between-schools within a year for curriculum revision or other instructional purposes. This could very well be the beginning of the broader 'item bank' where all the schools are branches. Teachers, under the guidance of HODs, could make meaningful and secured 'deposits' of their analyzed items. Whenever and wherever needed, these deposits could be drawn for specific uses - then replaced and replenished, whenever and wherever necessary. As PCs become linked to one another, quick 'electronic transfers' could also be made.

In conclusion, we would like to reiterate that assessment is very much part of teaching. This is even more so in a system in which assessment results are used for important decisions made on the students and the schools. More time and thoughts could be beneficially invested to develop teachers' assessment skills and relevant understanding, to enable teachers to discharge this aspect of their responsibility with greater productivity and efficiency, and to explore hitherto untried ways of dealing with assessment matters.

References

Nolen, S. B., Haladyna, T. M., & Haas, N. S. (1992). Uses and Abuses of Achievement Test Scores. *Educational Measurement: Issues and Practice*, Summer 1992, 9-15.

Schafer, W. (1991). Essential Assessment Skills in Professional Education of Teachers. *Educational Measurement: Issues and Practice*, Spring 1991, 3-12.

Stiggins, R. (1992). High Quality Classroom Assessment: What Does It Really Mean? *Educational Measurement: Issues and Practice*, Summer 1992, 35-39.

Soh Kay Cheng and **Poh Sui Hoi** are members of the staff of the Division of Psychological Studies, School of Education, National Institute of Education, Nanyang Technological University.

ESTHER TAN

Assessment in Guidance and Counselling

When we speak of assessment in guidance and counselling, we are referring to the evaluation or appraisal of pupils according to selected dimensions, parameters or categories of information to arrive at a description and understanding of the pupil's current status. At the developmental level, this kind of assessment helps to gather information about students which will help them develop self-awareness and make informed and meaningful decisions, as in the case of educational planning and career guidance. At the remedial level, the purpose of assessment is to ascertain the status and needs of the pupil so as to help the teacher plan appropriate and effective helping strategies, as in the case of counselling pupils with behaviourial problems.

Pupil assessment is a complex and dynamic process that depends on grasping the interaction of the individual's attributes with the environment. In recent years, testing instruments have made significant contributions in assessment by emphasizing the quantitative dimension of studying pupil behaviour. However, the limitations and imperfections of tests make it desirable that nontest techniques also be used.

Nontest Techniques in Assessment

Observation

Observation is the base for most nontesting appraisal techniques. Observing and recording descriptions of students have a number of important purposes for those who work with them. Observation may yield data that can challenge tentative hypotheses about the individual pupil and confirm others. It can also be used to evaluate the effectiveness of steps taken to facilitate the individual pupil's learning, development and adjustment.

There is no shortcut to mastering the skills of observation. To ensure that the behaviour observed is representative of the individual, a number of observations should be made in a variety of situations and at different times. The teacher observes the pupil at work and at play, noting carefully his behaviour when alone and his interaction pattern when he is with others. He can use frequency counts, namely, by recording how frequently a certain behaviour occurs, how serious these occurrences are and how long each episode lasts. He can also use event or time sampling methods to record behaviours in a systematic manner. Having collected data from different sources, the teacher is then ready to synthesize an accurate, representative and meaningful picture of the pupil.

At the remedial level, the purpose of assessment is to ascertain the status and needs of the pupil so as to help the teacher plan appropriate and effective helping strategies

The use of observation as an assessment technique, however, is not without problems. Unconscious biases in observation sometimes occur because observers fail to admit their own feelings and limitations or because they are unaware of them. Biased observers tend to attribute their own behaviourial tendencies to others. For example, an authoritarian teacher may view a student who disagrees with him as domineering or rebellious. Accurate observations require an ability to evaluate objectively what is being perceived as well as an awareness of one's own feelings and beliefs.

Misinterpretation of observed behaviour and inaccuracy in reporting are other problems that can destroy the usefulness of observations. Mental sets, interests, or expectancies often alter perceptions of behaviours or situations because each person tends to see the other's world as he or she has experienced it.

There are several things a teacher can do to improve his observation

skills. Practice will help increase the accuracy of observations, and practice in which two or more individuals observe simultaneously and compare results will be even more beneficial. Observe only one person at a time. Before observation takes place, determine what is to be observed. Spread observations over a period of time. If possible, try to record and summarise the observation immediately after it is completed to increase accuracy in recording.

The Interview

The interview is the most essential tool in counselling and serves important functions at the various stages of the helping process. In the early stage it is used for such general purposes as obtaining and giving information. Such fact-finding interviews differ from the counselling interview in that greater amount of control is exercised by the interviewer. Definite, usually pre-determined kinds of information are sought. The interviewer asks questions, probes for responses, encourages the interviewee to communicate fully and records the information revealed.

The counselling interview is much less structured. As the aim is to help the pupil develop self-understanding and a better grasp of the problem situation, the interviewer needs to engage in a great deal of active listening. He demonstrates empathy and listens to facts, ideas as well as feelings, being alert to both the verbal communication as well as the body language of the pupil. The skilful interviewer also masters the fine art of questioning, keeping in mind the relevance of the questions, the tone of the questions and the choice of words.

Anecdotal Records

Anecdotal records are brief informal reports by the teacher of an observation of a critical incident. They describe a sample of behaviour in a given situational context. The behaviour described may be positive or negative, but it is usually the pupil's behaviour that is described, not the teacher's interpretation of the behaviour.

Anecdotal records are useful in that they describe the behaviour of an individual in diverse situations, thus contributing to a fuller understanding of the individual's personality. They also supplement quantitative data and enrich interpretations of behaviour. However, anecdotal records can be valuable only to the extent that the observational description is accurate and comprehensive.

There are different opinions as to whether interpretation should be a part of anecdotal records. For practical purposes, it may be helpful to delay interpretation until several anecdotes have been obtained. This gives the teacher an opportunity to study patterns of behaviour and base his interpretation on a larger sample of pupil behaviour.

Self-Reports

Another informal method to find out about children is to have them talk or write about themselves. For pupils who are not so articulate or have a language barrier, the teacher can provide crayon and paper and encourage them to express their views in drawing. Ask them to draw their home, their family, their favourite games etc. and then encourage them to talk about their pictures. Children who cannot verbalise their feelings often reveal them through their drawings. With very young children, another approach is to engage them in play activities, not just to put them at ease, but also to observe their behaviour.

For older children with better language abilities, the teacher can let them write their own autobiographies which can be very revealing. To obtain even more specific information, the teacher can use the sentence completion technique e.g.

My greatest wish is _____
What I like most about school is ___
I worry when _____
In class I usually _____

How informative these self-reports are depends on the pupil's willingness to reveal himself and his ability to express himself. These self-reports may not be always accurate, but they do give some idea of how the child feels about himself and those around him.

Pupil-Data Questionnaires

Schools often use pupil-data questionnaires to obtain vital information about their students. Usually such questionnaires consist of items regarding the student's home, family, health, educational and vocational plans, out-of-school and in-school activities and study habits etc.

In administering pupil-data questionnaires, pupils should be told why the forms are important. The extent of confidentiality of the data obtained should also be discussed with them.

Teacher-Parent conferencing helps to establish rapport between the parent and the teacher leading to collaboration between home and school.

Teacher-Parent Conferencing

Teacher-Parent conferencing serves three purposes. Firstly, it helps to establish rapport between the parent and the teacher leading to collaboration between home and school. Secondly, interviewing the parent can reveal important background information such as the child's developmental history or the parent's child-rearing practices. Such information often sheds light on the pupil's behaviourial problems or learning difficulties in school. Thirdly, the interview allows the teacher a chance to interpret the school to the parent and solicit his understanding and support in the helping process.

The Case Study

The case study is a comprehensive method of collecting and summarizing data about an individual. The information contained in it is garnered from all available reliable sources: cumulative records, observations, interviews, autobiographies, data questionnaires and tests etc.

The intent of the case study is to obtain a thorough understanding of the troubled student so that appropriate intervention programmes can be planned. In order to be valuable, the case study must be written clearly, accurately and objectively with a minimum of personal bias in the interpretation. Irrelevant items, technical terms and generalizations unsupported by specific data should be avoided.

The Use of Tests in Guidance and Counselling

A test is generally a set of questions, problems, puzzles, symbols and exercises used to determine a person's ability, aptitude, knowledge, qualifications, interests and level of social adjustment. The kinds of tests used in guidance and counselling are usually psychological tests. These are constructed to assess a representative sample of an individual's behaviour from which the totality of that individual might be inferred. Hence, the more complex the behaviour to be assessed, the more difficult the testing task, and the response to that task. This explains why sometimes a battery of tests is used instead of one single test to help obtain a comprehensive and accurate assessment.

Functions of Tests

Tests can serve a number of functions in guidance and counselling. They can provide data to help the pupil increase self-understanding, self-acceptance and self-evaluation. In addition, test results can be used to challenge the pupil's perception of himself or herself and the world and can promote exploration in a number of areas. Tests can be used to predict a pupil's success in a specific course of study, job, career or other endeavour, as in the case of educational planning and career guidance. Tests can be used diagnostically to help a pupil understand better the skills and knowledge he or she possesses and to gain insight into areas that are below acceptable level. This can help the pupil identify his or her weak areas that require greater concentration or attention. Tests can also fulfil a monitoring function by helping the teacher or counsellor see what progress (or lack of progress) the pupil is making. Finally, tests can be used to evaluate the pupil's growth, the teacher or counsellor's success in the helping process or the achievement of certain set goals. In serving all these functions the test can be a predictive tool, a diagnostic aid, a monitoring device or an evaluative instrument.

Tests can provide data to help the pupil increase self-understanding, self-acceptance and self-evaluation.

Selection of Tests

Proper selection of tests requires knowledge of the types of tests available. One simple distinction is between standardized and nonstandardized tests. Standardized tests usually have a standard administrative process including instruction and specific time limit to ensure consistency in administration and measurement regardless of the administrator or the place of testing. There is also provision of scoring instructions or a scoring key designed to eliminate scorer errors on the test. Usually various normative data are made available with the test to allow for comparison with a wide variety of groups. A manual with technical testing data, such as validity and reliability, is also included.

Another distinction is between group tests and individual tests. Generally group tests require less formal and supervised preparation for administration than do individual tests. There are also distinctions between pencil-and-paper tests and performance tests. The latter require the use of objects and physical skills and generally provide more direct data for skill area judgment than do pencil-and-paper tests. The advantage of the former, however, is that they require less time and cost to administer.

A final distinction can be made between speed tests and power tests. Speed tests are designed to measure the examinee's speed of accomplishment while a power test measures the level of performance. Speed tests are usually so long that very few individuals can finish all the test items in the time provided. On the other hand, the test items in power tests are usually arranged in increasing order of difficulty and performance depends on the degree of successful completion of items.

Before a test is used in guidance and counselling, certain professional and ethical considerations need to be addressed.

Types of Tests

Checklists

The simplest type of tests used in guidance and counselling are checklists. These are easy to construct and simple to administer and are often used to yield a profile of a pupil's behaviour to help the teacher gain insight into the former's behaviour.

Behaviour checklists are a kind of self-report, when the pupil is asked to assess himself by providing a description of his own character in the form of a checklist in which he indicates by a tick the description that applies to him. Examples of such checklist items are:

- Often feeling sick
- Not interested in studies
- Get poor examination results
- Slow in making friends
- Afraid of failing in exams

The Mooney Problem Checklist developed by Ross Mooney (1950) and the Jesness Behaviour Checklist (1971) are two such checklists frequently used by school counsellors to identify problems for individual counselling. Checklists that are more appropriate for use with primary school children are Children's Problem Checklist (Schinka, 1985) and School Behaviour Checklist (Miller, 1977, 1981). If the pupil in question is a young child, the parent or class teacher can be asked to fill in a descriptive checklist instead of the child himself. In fact, the practice of having both the parent and class teacher assess the child can be very helpful as it can reflect the pupil's behaviour both in school and at home.

Rating Scales

The rating scale presents a list of descriptive words or phrases to be checked by the rater. These instruments are usually used to rate pupils on characteristics such as honesty, dependability, cooperativeness and self-reliance etc. The rating is usually made on a scale (normally 1-2-3-4-5) so as to allow adequate sensitivity.

Rating scales are usually considered to be more robust and valid than checklists as they give more precise information on the individual. Some examples of rating scales commonly used in guidance and counselling are Behaviour Rating Profile (Brown & Hammill 1978, 1983) and Behaviour Evaluation Scale (McCarney, Leigh and Cornbleet 1983), for both primary and secondary school pupils.

There are, however, a number of inherent problems in rating scales and checklists. For example, ambiguity error may occur when a term is understood in more than one way. In other instances, the tester may rate too high (leniency or generosity error), or too low (severity error). One way to minimise such errors is to have more than one rater or to combine scores.

Interest Inventories

The study of interests has been of major importance to career counsellors in their efforts to understand and assist pupils. The appraisal of vocational interests is usually accomplished by the use of a standardized inventory to obtain information relevant to educational and vocational decision making, either to confirm or to open up new possibilities.

The use of interest inventories in career guidance is gaining recognition and importance in Singapore schools. Amongst the inventories commonly used, John Holland's Self Directed Search is by far the most popular, being a self-administering and self-scoring test that could be administered in groups or as an individual test. This test yields a career interest profile of the individual along six dimensions, namely, realistic, investigative, artistic, social, enterprising and conventional. Other examples are Strong-Campbell Interest Inventory (Strong et al, 1981) and the Geist Picture Interest Inventory for young children (Geist, 1964). Indigenous career interest inventories are being developed by staff of the National Institute of Education, amongst them the Career Profile Inventory which will be available both as a pencil-and-paper test as well as a computer software package.

Personality Inventories

Two approaches are commonly used in assessing the psychological makeup of individuals. The first are personality questionnaires which belong to the pencil-and-paper, self-report variety. These tests are designed to measure such characteristics as emotional adjustment, social relations and the motivational aspect of behaviour. Examples of social traits assessed are ascendance-submission, introversion-extroversion and self-sufficiency.

In general, personality inventories are constructed on the assumption that human personality has a certain amount of stability and that over a range of similar situations, the same reactions will be elicited. Different individuals possess varying amounts of each trait. The more responses of a certain nature that examinees mark, the more likely they are to possess the trait being measured.

Some examples of popular inventories used in guidance and counselling are the Lewis Counselling Inventory (Lewis & Pumfrey, 1978) which is designed for adolescents and yields

five scores on relationships with teachers, family and peers as well as irritability and social confidence. The Behaviourial Academic Self-esteem (BASE) (Coopersmith & Gilberts, 1982), on the other hand, has five scales on student initiative, social attention, success/failure, social attraction and self confidence.

In an alternative approach, projective techniques are used in the appraisal of personality. A projective device places individuals in a situation in which they are asked to describe something, relate a story or respond to words or pictures. The underlying assumption is that in responding, individuals unwittingly reveal things about themselves as they often project their own feelings and problems into the situations. The individual's responses are then evaluated by the tester who is extensively trained in the interpretation of projective tests.

Ethical Issues in Psychological Testing

Before a test is used in guidance and counselling, certain professional and ethical considerations need to be addressed. The first concerns the competence and qualification of the tester. This is important because unwise and inappropriate use of psychological tests and inaccurate interpretation of test data may have harmful consequences. While some standardized tests can be administered by the classroom teacher with the help of manuals, teachers who do not have the required administration skills should refrain from using psychological tests, especially those where projective techniques are used in the scoring and interpretation of data. Other issues to be considered are examiner rapport and test anxiety. Research has shown that in any test situation, the level of text anxiety as well as the absence or presence of examiner rapport have significant effects on test results. Also, to safeguard against inaccurate interpretation and over-generalisation, one can use a battery of tests instead of just one test, or use tests alongside other assessment procedures and sources of gathering data about the pupil such as observation and interviews.

Conclusion

In developing the appraisal component of a guidance and counselling programme, one needs to think through the following:

1. How much tests and nontest data on pupils are really needed to help the pupils reach attainable goals?
2. To what extent should test data be directed towards remedial as opposed to developmental goals?
3. What provisions should be made for transmitting certain aspects of appraisal data to those who need to know such data?
4. Are the staff to be involved qualified to carry out testing on the pupils? If they require training in this area, how is the training to be provided?

Recognising the limitations of appraisal measures and bearing in mind the professional and ethical issues concerning the use of tests in guidance and counselling, teachers and counsellors should be highly selective and responsible in choosing and using assessment methods. Whatever the approach and whether test or nontest techniques are to be used, the ultimate goal should be for the benefit of the pupil.

References

Brown , L.L. & Hammill, D.D. (1978, 1983). Behaviour Rating Profile. Pro. Ed.

Coopersmith, S., & Gilberts, R. (1982). Behaviourial Academic Self-Esteem (BASE). California: Palo Alto, Consulting Psychologists Press.

Geist, H. (1964). Geist Picture Interest Inventory. California: Western Psychological Services.

Holland, J. L. (1983). Self Directed Search. California: Palo Alto, Consulting Psychologists Press.

Jesness, C.F. (1970). Jesness Behaviour Checklist. California: Palo Alto, Consulting Psychologists Press.

Lewis, D. G. & Pumfrey, P.D. (1978). The Lewis Counselling Inventory. Windsor, Berks, NFER.

McCarney, S.B., Leigh, J.E., & Cornbleet, J. (1983). Behaviour Evaluation Scale. Columbia: Educational Services.

Miller, L.C. (1977). School Behaviour Checklist. Los Angeles: Western Psychological Services.

Mooney, R.L., & Gordon, L.V. (1950). Mooney Problem Checklist. San Antonio: The Psychological Corporation.

Schinka, J.A. (1985). Children's Problem Checklist. Odessa: Psychological Assessment Resources.

Strong, E.K., Campbell, D.P., & Hansen, J.C. (1981). Strong-Campbell Interest Inventory. California: Stanford University Press.

Esther Tan is Senior Lecturer and Head, Division of Psychological Studies, National Institute of Education, Nanyang Technological University, Singapore.

GRANT WIGGINS

Creating Tests Worth Taking

The Director of Research at CLASS provides questions, criteria, and suggestions for test designers who want to engage students as well as evaluate their performance.

Should a test be enticing? I think so. And should tests more often be authentic simulations of how knowledge is tested in adult work and civic settings? Many of us believe so. "Performance assessment" calls upon test makers to be creative designers then, not just technicians.

In performance assessment the design issues resemble those facing the architect. There is ground to be covered (the syllabus), there are the logistics of making the design fit the site (making large-scale assessment work in the school), and there are building codes (psychometric norms) and town elders (school board members and district testing directors) to worry about. But designers have typically avoided another, more basic obligation: the need to serve the users - in this case, students and teachers. The clients must "own" the design: form must follow function. The more the tasks (like the house) fit seamlessly with both the environment and the client's aspirations, the better the design and the result.

In this article, I offer some proven design tips, tools, and criteria for fashioning assessment tasks that are more enticing, feasible, and defensible - tests worth taking.

Questions and Criteria

Designers of performance assessments should use the following key questions as a tool to guide the design process:

- What kinds of essential tasks, achievements, habits of mind, or other valued "masteries" are falling through the cracks of conventional tests?
- What are the core performances, roles, or situations that all students should encounter and be expected to master?
- What are the most salient and insightful discriminators in judging actual performances?
- What does genuine mastery of each proposed assessment task look like? Do we have credible and appropriate exemplars to anchor our scoring system? Have we justified standards so they are more than local norms?
- Are the test's necessary constraints - imposed on help available from others, access to resources, time to revise, test secrecy, prior knowledge of standards - authentic?
- Do our assessment tasks have sufficient depth and breadth to allow valid generalizations about overall student competence?
- Have we ensured that the test will not be corrupted by well-intentioned judges of student work?
- Who are the audiences for assessment information, and how should assessment be designed, conducted, and reported to accommodate the needs of each audience? When are audit-tests appropriate and inappropriate?

These questions can be summarised and reframed to produce eight basic design criteria:

1. Assessment tasks should be, whenever possible, authentic and meaningful - worth mastering.
2. The set of tasks should be a valid sample from which apt generalizations about overall performance of complex capacities can be made.
3. The scoring criteria should be authentic, with points awarded or taken off for essential successes and errors, not for what is easy to count or observe.
4. The performance standards that anchor the scoring should be genuine benchmarks, not arbitrary cut scores

Typical tests, even demanding ones, tend to overassess student "knowledge" and underassess student "know-how with knowledge"

or provincial school norms.

5. The context of the problems should be rich, realistic, and enticing - with the inevitable constraints on access on time, resources, and advance knowledge of the tasks and standards appropriately minimized.

6. The tasks should be validated.

7. The scoring should be feasible and reliable.

8. Assessment results should be reported and used so that *all* customers for the data are satisfied.

The suggestions and observations that follow offer further assistance to would-be designers.

Choosing What to Test

Choose exit outcomes or areas of the curriculum that now tend to fall through the cracks of conventional testing. Typical tests, even demanding ones, tend to overassess student "knowledge" and underassess student "know-how with knowledge" - that is, intellectual performance. Auditing local tests with Bloom's taxonomy as criteria, for example, shows that synthesis is infrequently assessed at present, and is *inherently resistant* to assessment by multiple-choice tests because it requires "production of a unique communication" that bears the stamp of the student[1].

Faculties should also consider their institutional "customers." What kinds of tasks must our *former* students master? Here, for example, is a question from a freshman final exam in European history at a prestigious college; it suggests how even our better students are often ill-prepared for real intellectual tasks:

> Imagine yourself Karl Marx, living half a century later. Write a brief evaluation of the programs of the Fabian socialists and the American reformers such as T. Roosevelt to present to the Socialist International.

Think of the knowledge to be tested as a tool for fashioning a performance or product. Successful task design requires making the essential material of a course a *necessary means* to a successful performance *end.* Example: a 5th grade teacher assesses geography knowledge by having his students devise a complete itinerary, map, and travel packet for their favorite rock group's world tour, within certain budget, logistical, cultural and demographic restrictions.

Another example: students are asked to design a museum exhibit around a theme studied in a history course, selecting from many real or facsimile artifacts; required to justify what is both included and excluded in the exhibit; and must seek funding from a "foundation" of teachers and peers for the exhibit.

We want to know: Can the student use knowledge and resources effectively, *to achieve a desired effect?* This is the question Bloom and his colleagues argued was at the heart of synthesis. These tasks should only be judged well done to the extent that the content is well used.

Designing the Tasks

Contextualize the task. The aim is to invent an authentic simulation, and like all simulations, case studies, or experiential exercises, the task must be rich in contextual detail. A context is rich if it supports multiple approaches, styles, and solutions and requires good judgments in achieving an effective result. One must please a real audience, make a design actually work, or achieve an aesthetic effect that causes pride or dismay in the result.

The test may be a contrivance, but it needn't *feel* like one.[2] Consider professional training and testing. Doctors and pilots in training confront situations that replicate the challenges to be later faced. Business and law students learn by the case method, fully immersed in the facts of real past cases. A context is realistic to the extent that we so accept the premises, constraints, and "feel" of the challenge that our desire to master it makes us lose sight of the extrinsic factors and motives at stake - namely that someone is evaluating us. In just this way, for example, putting out a school newspaper for a journalism course doesn't feel contrived.

Here's an example of how a teacher's attempt to design a performance task evolved as a concern for context was introduced. The original task, in a global studies course, required students to design a trip to China or Japan. But what kind of trip? For what customers? With what constraints of budget or time? The teacher then refined the task so that each student had a $10,000 budget for designing a month-long, cultural-exchange trip for students their age. Still, the purpose is too abstract. What must the tour designers accomplish? Are they trying to design a tour in the abstract or really attract tour-takers? The students were finally charged to be travel agents who develop an extensive brochure and research the cost and logistical information using a computer reservations system.

There is no such thing as performance-in-general. To understand what *kind* and *precision* of answer fits the problem at hand, the student needs contextual detail; it clarifies the required result, hence the criteria and standards. Too many measurement tasks have an acceptable margin of error that is arbitrary. Are we measuring body temperature or roasts in the oven? It matters. The task's standard of performance (desired precision or quality of product) should be apparent. In fact,

All real-world performers know the target and the standards, not just their task in advance.

an important oversight by the global studies teacher was her failure to give the students model tour brochures.[3]

Aim to design "meaningful" tasks - not the same as "immediately relevant or practical" tasks. An assessment task will be meaningful to the extent that it provokes thought and thus engages the student's interest. But a task can be engaging without being of apparent, immediate usefulness. Whether it be mysteries, debates, mock trials, putting on plays - or, for that matter, Nintendo - students clearly respond to "irrelevant" but real challenges. What do such tasks have in common? Designers need to conduct better empirical studies to discover the tasks that tap those twin intellectual needs: *our urge for efficacy and our need for meaningful connections.*

This caution about meaning vs. relevance is particularly warranted to avoid turning important theoretical problems into crude utilitarian ones. Many genuine problems do not have obvious practical value, but they nonetheless evoke interest and provide insight into student abilities. Consider two such problems, one in geometry and one in history/English:

Problem 1

We all know the Pythagorean theorem: $A^2 + B^2 = C^2$; but does it have to be a square that we draw on each leg? Suppose we drew the same shape on each leg; would the areas on A and B add up to the area on C? Find other shapes that make the equation work, too, and try to derive a more general formula of the theorem.[4]

Problem 2

You and your colleagues (groups of 3 or 4) have been asked to submit a proposal to write a U.S. history textbook for middle school students. The publishers demand two things: that the book hit the most important things, and that it be interesting to students. Because of your expertise on the 18th century, you will provide them a draft chapter on the 18th century, up to but not including the Revolution, and "field-tested" on some middle school students. They also ask that you fill in an "important" chart with your response to these questions: (1) Which event, person, or idea is most important in this time period, and why? (2) Which of three sources of history - ideas, people, events - is most important? You will be expected to justify your choices of "most important" and to demonstrate that the target population will likely be interested in your book.

Design performances, not drills. A test of many items (a drill) is not a test of knowledge in use. "Performance" is not just doing simplistic tasks that cue us for the desired bit of knowledge. It entails "putting it all together" with good judgment; good judgment cannot be tested through isolated, pat drills. As one teacher put it to me a few years ago: "The trouble with kids today is that they don't know what to do when they don't know what to do." She is right - and a prime reason is that tests rarely put students in an authentic performance situation, where *thinking*, not just an obvious bit of knowledge, is required.

The designer's aim, then, is to avoid inventing a new round of (this time, hands-on) isolated items. Rather, we should consider the difference between drilled ability vs. performance ability and ask: *What is the equivalent of the game or recital in each subject matter?* What does the "doing" of mathematics, history, science, art, language use, and so forth, look and feel like in context. What are the projects and other kinds of synthesizing tasks performed all the time by professionals, consumers, or citizens that can be adapted to school use?

Such tasks are always "higher-order," and we would do well to use Lauren Resnick's criteria in our search for better-designed assessments. Higher-order thinking

- is *nonalgorithmic* - that is, the path of action is not fully specified in advance;
- is *complex*, with the total path not visible from any single vantage point;
- *often yields multiple solutions*, each with costs and benefits;
- involves *nuanced judgments* and interpretations;
- involves the *application of multiple criteria*, which sometimes conflict with one another;
- often involves *uncertainty*, because not everything that bears on the task is known;
- involves *self-regulation* of the thinking process, rather than coaching at each step;
- involves *imposing meaning*, finding structure in apparent disorder;
- is *effortful*, with considerable mental work involved.[5]

It may help to think of this problem as the search for larger, more interrelated but complex chunks of content to build tasks around. What, for example, might be 8 to 10 important performance tasks in a subject that effectively and efficiently "map" the essential content? Vocational programs usually grapple well with this problem by casting the course objectives as a set of increasingly complex tasks to be mastered, in which the student in the last task(s) must literally put it all together, for example, build a house in carpentry.

Refine the tasks you design by building them backwards from the models and scoring criteria. A complex task is not a vague task, with the objective or

specifications unknown. All real-world performers know the target and the standards, not just their task in advance; such knowledge guides their training and rehearsals. Students should never have to wonder "Is this right?" "Am I finished?" "How am I doing?" "Is this what you want?" In a "real" problem the task is ill-structured but well-defined; the goal, specifications, or desired effect is known, but it is not obvious how to meet it. Knowing the *requirements of task mastery* - the "specs" - means the student must be habituated by testing to think of mastery as control over the *knowable* essentials, not as calculated cramming and good guesses. This requires providing the student with scoring criteria and models of excellent performance or productions as part of instruction. (Think of diving and debate.) Such practice is the norm throughout Carleton, Ontario, where students work from "exemplar booklets" to practice grading student work - in the same way now reserved for judges in our assessments.

"What does mastery at the task look like? What will we be able to properly infer from the collected student work?" These become the *key* questions to ask in the challenge of taking a basic idea and making a valid performance-assessment task out of it (as opposed to an instructional task). The questions properly focus on judging anticipated results and move away from design that produces merely pleasant or interesting work.

Scoring Considerations

Score what is most important for doing an effective job, not what is easiest to score. The scoring rubrics should represent generalizations about the traits found in an array of actual performances. But too often we resort to scoring what is easiest - or least controversial - to observe. A fine task can be rendered inauthentic by such bogus criteria.

Two key questions for setting up a scoring system therefore are: "What are the most salient characteristics of each level or quality of response?" and "What are the errors that are most *justifiable* for use in lowering a score?" Obvious successes and errors (such as those that relate to spelling or computation) are not necessarily the most accurate indicators of mastery or its absence.[6] Too many essay scoring systems reward students for including merely *more* arguments or examples; quantity is not quality, and we teach a bad lesson by such scoring practices.

When possible, scoring criteria should rely on descriptive language, not evaluative and/or comparative language such as "excellent" or "fair." Judges should know specifically where in performance to look and what to look for. The ACTFL foreign language proficiency guidelines and the Victoria, Australia, "Literacy Profiles" are perhaps the best examples available of such empirically grounded criteria.[7] Teachers may also want to have students analyze a task and help devise the scoring system. This builds ownership of the evaluation, makes it clear that judgments need not be arbitrary, and makes it possible to hold students to higher standards because criteria are clear and reasonable.

"Benchmark" the standards for performance to ensure that your scoring standards are wisely chosen and suited to wider-world or next-level demands. Standard-setting for performance involves selecting exemplary samples of performance or production. The challenge is to avoid using local age-grade norms; the solution is to equate our exit-level standards to wider-world entry-level standards at desired colleges or professions. That advice, of course, begs a more fundamental question: Whose view of excellence should count? It is at least prudent to equate local standards of scoring to some credible wider-world or next-level standard - something routinely done in the performing arts, athletics, and vocational education.[8] *And, every so often, refer to next-level standards when scoring the work of younger students.* (I believe Illinois was the first state to assess both 6th and 8th grade writing samples against 8th grade exemplars, for instance).

Assessment design is like software design: one can never accurately and fully anticipate the naive user's response.

Administering the Assessments

Since constraints always exist in testing, make them as authentic as possible. The question is not "Should there be constraints in testing?" but rather "When are constraints authentic, and when are they inauthentic?" It is often a matter of degree, but the principle needs to be maintained and defended.

Constraints facing the designer of authentic assessment tasks typically involve access or restrictions to the following resources: (1) time (including time to prepare, rethink and revise), (2) reference material, (3) other people (including access to peers, experts, the test designer, and/or the judge), and (4) prior knowledge of the tasks and how they will be assessed (the issue of test security). The question then becomes: What are *appropriate* limits on the availability of these resources?

Traditional testing, because it involves indirect proxies for performance, requires numerous inauthentic constraints to preserve validity. The validity of most multiple-choice tests, for example, is compromised if questions are known in advance or if reference materials can be consulted during the test. These habits of ad-

Good teaching is inseparable from good assessing.

ministration run deep; they seem obviously required. But what of the validity issues raised by denying students access to basic resources? Just what is being tested when the student cannot predict the historical periods or books that will be assessed, or cannot consult resources while writing?

We need not keep textbooks and other materials from students if the task is genuinely authentic. For example, in many of Connecticut's performance tasks in mathematics, the key formulas are given to the student as background to the problem. And why not allow the student to bring notes to the exam? A physics teacher I know allows students to bring an index card to the exam with anything on it; the card often reveals more about the student's knowledge than the exam answers!

Too little time for performing is not always the key issue either. Is the limiting of the test to *one sitting* authentic? If writing is indeed revision, for example, why not allow writing assessment to occur over three days, with each draft graded? Many districts now do so, including Jefferson County, Kentucky, and Cherry Creek, Colorado[8].

I am not arguing that the student should have unlimited time and access in testing.[10] Let us ask: What kinds of constraints authentically simulate or replicate the constraints and opportunities facing the performer in context? What kinds of constraints tend to bring out the best in apprentice performers and producers?

Develop a written, thorough protocol that details how the task should be administered - especially so judges will know the proper limits of their interventions to student acts, comments, or questions. It is incredibly easy to invalidate performance assessment by varying the instructions, the amount of assistance provided, and the depth of responses given to inevitable student questions. Determining beforehand what is acceptable response and intervention by adults is essential; test administrators must receive standard verbal responses for delicate situations, confusions, or problems that arise.

And don't forget that kids can do the darndest things with directions that aren't thought through. In a hands-on science experiment that asked whether "the sun" heated up different colored liquids at different rates, a student did not use the heat lamp provided, moved all his equipment to the window, saw it was a cloudy day, and wrote "no."

Make the tasks maximally self-sustaining and the record-keeping obligation mostly the student's. Many educators who have never seen large-scale performance assessment cannot fathom how all students can be efficiently and effectively assessed. But they assume that the teacher will have to guide activity every step of the way and record massive amounts of information simultaneously. Thoughtful preparation, designed to make the assessment self-running, frees the teacher to be a perceptive

Horace's School: Redesigning the American High School.

Theodore Sizer
Boston:
Houghton Mifflin Company, 1992/

This book is a valuable tool for a school in the midst of a major assessment/restructuring process. Presented as an extended case study entering around the fictional teacher, Horace Smith, the book follows Horace through a series of restructuring committee meetings he is chairing at Franklin High School.

The meetings accurately capture the blend of tedium and excitement characteristic of the committee process. Flowing from the discussions of what it means to be well educated and how to best provide this education are several examples of "exhibitions." These are the means whereby students demonstrate their understanding of ideas and skills underlying the school's newly devised program. The exhibitions provide readers Sizer's best examples of performance assessments for high school students.

Interspersed with Sizer's commentary is his narrative. It is here I found him at his best. The chapter "Policy and Power" is as cogent and heartfelt a statement about reform as I have read.

The book draws from years of research and the author's work with the Coalition of Essential Schools. Sizer delineates the Coalition's "nine common principles," which recognize there is no one way for a good school to look or proceed. Likewise, there are no shortcuts in the restructuring process, especially as it seeks to challenge the underlying principles of our current schools. This book sheds light on the reform process and helps clarify the challenge. The rest is up to us.

Available from Houghton Mifflin Company, Two Park St., Boston, MA 02108, for $19.95 (paperback).

- Reviewed by Stephen Garger, University of Portland, Portland, Oregon.

judge.

Creating a Tool Kit

Develop a districtwide "tool kit" of exemplary tasks, task templates, and design criteria for assessment tasks. Not all of us are good designers, but why should we have to be? Teachers can help their colleagues by providing a sampler of tasks and task templates. Kentucky has done this at the statewide level, providing dozens of tasks and task ideas to teachers as part of the new state performance-based assessment system. We should consider including not only current examples of model assessment tasks, but traditional performance-based challenges such as debates, treasure hunts, mysteries, design competitions, historical reenactments, science fairs, Odyssey of the Mind tasks, Scout Merit Badges, student-run banks and stores, and so forth.

The mathematics performance assessment team of the Connecticut Department of Education has identified the following types of problems as central to its work:

- Given data on graphs, write a story that represents the data or graph.
- Given headlines or claims with background data, explain whether or not the claims are reasonable.
- Given student work containing common errors, write a response to the student.
- Given equations or number facts, write a problem that the equations or facts could solve.
- Given trends or sample data, make and justify predictions.
- Given consumer- or job-related buying, selling, or measuring situations, solve a problem.
- Given multiple or competing interpretations of given data, justify each interpretation.

Job roles provide ample opportunities for task designers to create simulations. Here are some suggestions:

- Museum curator: design museum exhibits; compete for "grant" money.
- Engineer or surveyor: bid and meet specs for largest-volume oil container; build a working roller coaster; map or survey a region around school or in the building.
- Ad agency director: design advertising campaign, book jackets or blurbs for books read in class.
- Psychologist/sociologist: conduct surveys, perform statistical analyses, graph results, write newspaper articles on the meaning of the results.
- Archaeologist: determine the culture or time frame of a mystery artifact or person.
- Policy analyst: predict the future in a country being studied.
- Product designer: conduct research, design ad campaign, present proposal to panel.
- Job interviewee: present portfolio and try to get "hired" for a specific job related to skills of current course (interview conducted by other students or teacher).
- Expert witness to Congress: testify on behalf of or against advertising claims, regulation of children's TV, or current policy issue.
- Commercial designer: Propose artwork for public buildings.

Piloting and Reporting

Always pilot some or all of the test. Assessment design is like software design: one can *never* accurately and fully anticipate the naive user's response. A first design may not fit the purpose or maximally evoke the desired knowledge; a prompt might result in irrelevant responses that are nonetheless appropriate or reasonable to the student; the logistical constraints of a context can turn out to be more daunting than anticipated; the judges may be too self-interested in the results or insufficiently trained. A pilot is the only way to find out, even if it involves only a tiny sample of performers. And the de-bugging requires a *naive* "guinea pig" - a teacher from a different subject or a few students - if the hidden problems in the goal, directions, or procedures are to be found.

You are what you report: Make sure that your report cards, transcripts, and district accountability reports relate achievement and progress to essential performance tasks and exit-level standards. Few transcripts reflect achievement in reference to outcomes. They tend to certify that tests were passed on each isolated packet of content instead of documenting what the student can do and to what level of performance. Further, a one-shot test cannot validly assess many important capacities, as the phrases "habits of mind" or "consistency of performance" suggest. Grading and reporting thus need to move toward scoring that provides a "progress" measure - that is, work scored against exit-level performance standards. And no worthy performance is reducible to one aggregate score. Every student ought to have the equivalent of a baseball card - many different kinds of abilities measured and a brief narrative report - if we are seriously interested in accurately documenting and improving complex performance.

Assessment's Role in School Reform

An underlying premise of this kind

of assessment reveals why I believe that assessment reform is the Trojan horse of real school reform. We badly need better definitions of mastery or *understanding* to guide assessment design, curriculum design, and teacher job descriptions and performance appraisal. Circling "correct" answer to problems only test makers care about is not "knowing," nor is it the aim of teaching. Authentic tests provide a stimulating challenge instead of an onerous obligation.

Perhaps more important for school restructuring is the need to build local educator capacity and interest in quality assessment.[11] Genuine faculty empowerment is impossible without deep ownership of local standards and measures. Farming all these problems out to distant "experts" is a grave mistake - one rarely made in any other country. Good teaching is inseparable from good assessing. It may well be that experts can design more rigorous tests, and that correlational/predictive validities exist in standardized tests. But schooling we can be proud of and held genuinely accountable for demands more locally useful, authentic, and enticing assessments.

Notes

[1]Bloom, (1954) pp. 163, 175. Serious would-be test designers would do well to reread the *text* of the taxonomy, not just the Appendix/list, as well as the follow-up handbook developed by Bloom, Madaus and Hastings, (1981).

[2]"The student should [have] freedom from excessive tension ... be made to feel that the product of his efforts need not conform to the views of the instructor ... [and] have considerable freedom of activity ... [including] freedom to determine the materials or other elements that go into the final product." In Bloom, (1954), p. 173.

[3]See Linn, Baker, and Dunbar, (1991), for further discussion of validity design issues.

[4]I have watched half a dozen classes immerse themselves in this problem and beg to continue when time ran out.

[5]From Resnick (1987).

[6]Describing key errors and using them in the rubric is a *very different* matter than building them into test answers as "distractors".

[7]A related issue that emerges in designing rubrics (and thus far unaddressed by measurement experts) is the difference between the degree of difficulty of the task and the desired quality of the performance - a distinction made in New York's music performance assessments.

[8]See Wiggins (1991).

[9]Yes, yes, I know the issue is *really* one of cheating. Let the teacher "sign off" on the papers, then, certify authorship, as they do in Australia and now in Vermont.

[10]Though many New York State tests do allow the student what amounts to unlimited time - all day - given the shortness of the test. And certifiably learning disabled students are allowed unlimited time on the SATs as well as many state achievement tests.

[11]See Stiggins (1991).

Reference and Readings

Bloom, B.S., ed. (1954). *Taxonomy of Educational Objectives: The Classification of Educational Goals; Handbook I: The Cognitive Domain.* New York: Longman Publishers.

Bloom, B.S., G.F. Madaus, and J.T. Hastings. (1981). *Evaluation to Improve Learning.* New York: McGraw-Hill. [A major revision of *Handbook on Formative and Summative Evaluation of Student Learning.* (1971). McGraw-Hill].

Linn, R.L., E.L. Baker, and S.B. Dunbar. (November 1991). "Complex, Performance-based Assessment: Expectations and Validation Criteria." *Educational Researcher* 20, 8: 15-21.

Mitchell, R. (1992). *Testing for Learning: How New Approaches to Evaluation Can Improve American Schools.* New York: The Free Press.

National Council of Teachers of Mathematics. (Forthcoming). *Mathematics Assessment: Myths, Models, Good Questions, and Practical Suggestions.* Reston, Va.: NCTM.

Resnick, L. (1987). *Education and Learning to Think.* Washington, D.C.: National Academy Press.

Schwartz, J.L., and K.A. Viator, eds. (1990). "The Prices of Secrecy: The Social, Intellectual, and Psychological Costs of Testing in America." A report to the Ford Foundation, Education Technology Center, Cambridge, Mass.: Harvard Graduate School of Education.

Stiggins, R. (March 1991). "Assessment Literacy." *Phi Delta Kappan* 72, 7: 534-539.

Victoria Ministry of Education. (1990). *Literacy Profiles: Handbook.* Victoria, Australia. [Distributed in the U.S. by TASA, Brewster, N.Y.].

Wiggins, G. (February 1991). "Standards, Not Standardization: Evoking Quality Student Work." *Educational Leadership* 48, 5:18-25.

Wiggins, G. (June 1990). "Finding Time." *Basic Education* 34, 10.

Wiggins, G. (May 1989). "A True Test: Toward More Authentic and Equitable Assessment." *Phi Delta Kappan* 70, 9: 703-713.

Wiggins, G. (April 1989). "Teaching to the (Authentic) Test." *Educational Leadership* 46, 7: 41-47.

Grant Wiggins is the Director of Research and Programs for the Center on Learning, Assessment and School Structure (CLASS), 39 Main St., Geneseo, NY 14454. These design ideas derive from the work of CLASS in Frederick County, Md.; South Orange-Maplewood, N.J.; Urbandale Iowa.; and with faculties in Monroe, Orange, Suffolk, Ulster and Wayne counties, N.Y.

Reprinted with permission from Educational Leadership (May 1992)

MARY E DIEZ AND C JEAN MOON

What Do We Want Students to Know? ... and Other Important Questions

Answers to four questions will guide schools as they develop new models of curriculum and assessment

The successful use of new assessment strategies requires new assumptions about teaching, learning and assessment. A traditional focus on the delivery of information and the standardization of the circumstances of that delivery needs to give way to an emphasis on the development of learners.

Based on nearly 20 years of experience of the faculty of Alverno College, as well as our own work in assisting schools and districts with the change process, we suggest four questions that will help schools and teachers guide the development of new models of curriculum and assessment.

What Do We Want Students to Know and Be Able to Do?

In past practice, teachers and schools have often focused almost solely on content. The goals for an English course, for example, would be expressed in terms of the literary genres students will study. Asking what we want students to know and be able to do forces a more expansive look at curriculum goals and also raises the issue of relevance: To what end will they study literature? What personal abilities (thinking, empathy for others, self-expression) can be developed through interaction with works of literature?

We may, indeed, want to specify that students gain certain understandings or skills (as is done in the "required figures" part of a figure skating competition). But mightn't we also think about multiple modes of demonstration of the outcomes as well (as in the skater's "free program")?

The California social studies curriculum provides an example of an integrated goal, where content and performance come together: Students will demonstrate empathy for different periods of history[1]. The "required figures" (for example, the analysis of specific social, economic, political, and religious events and relationships) will depend upon what periods of history within which we choose to embed the development of this ability. But we could learn a great deal about the student by allowing for "free program" expression in a variety of student-selected demonstrations (written products, group projects, in-

Asking what we want students to know and be able to do forces a more expansive look at curriculum goals

tegrated art and writing, and so on).

What Will Count as Acceptable Performance?

Determining criteria for satisfactory performance may be the most difficult aspect of assessment. It requires us to back up and ask, "What would a student do if he or she had mastered a specific ability?" This means that we must examine what is at the heart of any and all competent performance, without being tied to the specifics of a particular performance.

In determining criteria for academic assessment, for example, we must look at the situation and set of directions that elicit the performance. For example, if we created an assessment in an American history course to measure a student's ability to demonstrate empathy for different periods of history, we might select the Civil War as a context and offer the following alternative directions (inviting students to suggest others):

1. Write a diary as though you were the mother of two sons during the Civil War, one fighting for the South and one for the North. Attach a statement about what you think was hardest for the mother.
2. Create a play about a family in the Civil War, where the action revolves around the decision of a member of the family to join the army. Attach a commentary about how the members of this family are like or unlike families you know.
3. Create a chart of aspects of the Civil War that affected families. Compare the experiences of families during the recent Persian Gulf War.

A second type of criteria, to assess quality, would relate to how well students demonstrate the goals; through some comparison with their own experience, how well do students connect an analysis of aspects of this period of history to how they themselves might feel had they lived at that time?

Performance criteria need to be general enough to allow students to practice what they will be judged on, without memorizing specific answers. They also need to be applicable to other periods of history, so that we can assess students' use of the ability in other contexts. Using both types of criteria to assess students in any one of the three tasks about the Civil War should show us whether students are seeing the period in historical perspective, as well as making links to their own experience.

How Can We Assure Expert Judgments?

The ability to use the criteria to determine the quality of students' work is what we mean by expert judgment, and it is far from the subjective process some fear. Because the criteria are known in advance to the students (and teachers use them to design learning experiences that lead up to the appraisal), assessment becomes a matter of gathering evidence in the student's performance to support a judgment whether each criterion is met.

Some criteria we might identify to assess students' ability to perform any of the three tasks about the Civil War noted above might include:

1. *Accurately uses information from the historical period (no evidence of anachronisms).* When students use information to create a picture of life in a specific historical period, the teacher can see the depth of their understanding; obviously, the presence of 20th century devices (televisions, fax machines), for example, would reveal problems in a student's grasp of the period.

2. *Uses sufficient detail to create a sense of what it was like for people who lived at the time under study.* This criterion calls upon the teacher's and students' sense of "how much is enough?" The teacher should talk through the need for the performance to satisfy an audience's need: Who will read what the student produces? What context-setting information will the audience need? How much description and how many

Determining criteria for satisfactory performance may be the most difficult aspect of assessment. It requires us to back up and ask, "What would a student do if he or she had mastered a specific ability?"

examples will be enough to paint a vivid picture for the reader? Having the teacher and students explore beforehand the meaning of "sufficient detail" can be an effective way to make the criterion clear to students.

3. *Draws out relationships or comparisons between that period of history and the present.* The criterion addresses critical thinking needed to make relationships, draw inferences, and engage in analysis. Both the teacher and students should examine the appropriateness and accuracy of the comparisons. Are these only the most obvious? Are they the most significant? Understanding and making relationships is not a skill that needs to wait until middle or high school. Elementary school children can learn to identify similarities and differences.

4. *Uses affective language in dealing with the experiences of people - in history and today.* This criterion requires preparation, as do the others, in the learning experiences that build up to the assessment. Our traditional testing practices have not emphasized affective goals. What better way to begin to make history meaningful than to see it as affecting the way people feel about their lives? Both the teacher and students need to ask questions like: Does the affective language capture what it might feel like to live in a period of war, given the circumstances of the time? Does the student link the way people in that period might have felt with his or her own feelings in a similar experience?

By being explicit and open about the criteria and giving students many examples of excellent work, we give them guidelines for improvement.

How Can We Provide Feedback?

Because the types of performances we've described do not reflect single solutions or memorized facts, they enhance opportunities for discussion with students and their parents. How well is a student learning to make relationships, not only in history, but in science, mathematics, and literature as well? Is he or she using this skill at home in the practical decisions of daily life? How can the student, teacher, and parents strengthen this ability to make relationships? Used in this way, the criteria become guides for learning, not an end point to the learning process.

Performance criteria need to be general enough to allow students to practice what they will be judged on, without memorizing specific answers.

The value of teachers and parents communicating productively is evident in the comments of a 1st grader teacher in Milwaukee who is developing assessment criteria in mathematics:

> For the first time, I had a lot of things to talk about when discussing mathematics. I was able to present parents with many examples of their child's development in terms of specific assessment criteria. My knowledge of the criteria provided me with the language necessary to present my assessment of students to parents.

Linking Learning and Assessment

By answering these four questions, we believe that schools and teachers can connect teaching, learning, and assessment in a meaningful way. With explicit goals and standards, teachers' expert judgment becomes a vehicle for informing learners about their progress toward specified goals and guiding them toward improvement.

[1] F. Alexander and C. Crabtree. (1988). "California's New History-Social Science Curriculum Promises Richness and Depth." *Educational Leadership* 46, 1: 10-13.

Mary E Diez is Chairperson, Education Division, Alverno College, 3401 S. 39 St., P.O. Box 343922, Milwaukee, WI 53234-3922. **C Jean Moon** is Director, Center for Math, Science, and Technology in Education, Lesley College, 29 Everett St., Cambridge, MA 02138-2790.

Reprinted with permission from Educational Leadership (May 1992).

WALTER SZETELA AND CYNTHIA NICOL

Evaluating Problem Solving in Mathematics

Effective assessment of problem solving in math requires more than a look at the answers students give. Teachers need to analyze their processes and get students to communicate their thinking.

In its *Curriculum and Evaluation Standards for School Mathematics*, the National Council of Teachers of Mathematics expanded the goals it developed in 1980 for promoting problem solving as a curricular focus (NCTM 1989). The first three standards - Mathematics as Problem Solving, Mathematics as Reasoning, and Mathematics as Communication - show a shift from emphasis on rules and routine problem solving dominated by teacher talk and passive learning, to active student participation, in which reasoning and communicating are stressed.

These efforts are admirable, but they create new challenges, especially in assessment of these higher-level skills. Problem solving requires considerable thinking, but even when students are able, they are not inclined to communicate their thinking. Without such communication, how can we reliably assess students' efforts to solve problems? Before discussing how to improve communication and assessment, it is useful to clarify the notion of a problem and problem solving.

The Nature of Problems and Problem Solving

Problem solving is the process of confronting a novel situation, formulating connections between given facts, identifying the goal, and exploring possible strategies for reaching the goal. A problem, then, is a situation in which the individual initially does not know any algorithm or procedure that will guarantee solution of the problem, but the individual desires to solve it.

Success in problem solving depends upon metacognitive processes, as described by Garofalo and Lester (1985). The following list summarizes the typical sequence of actions for successful problem solving:

1. Obtain an appropriate representation of the problem situation.
2. Consider potentially appropriate strategies.
3. Select and implement a promising solution strategy.
4. Monitor the implementation with respect to problem conditions and goals.
5. Obtain and communicate the desired goals.
6. Evaluate the adequacy and reasonableness of the solution.
7. If the solution is judged faulty or inadequate, refine the problem representation and proceed with a new strategy or search for procedural or conceptual errors.

These metacognitive processes are difficult to assess, but assessment can be expedited by creating problem situations that facilitate students' communication of their thinking.

Difficulties in Assessment of Problem-Solving Performance

The difficulty of assessing complex processes necessary for solving

Problem solving requires considerable thinking, but even when students are able, they are not inclined to communicate their thinking.

FIGURE 1

ANALYTIC SCALE FOR PROBLEM SOLVING

Understanding the problem

0 No attempt
1 Completely misinterprets the problem
2 Misinterprets major part of the problem
3 Misinterprets minor part of the problem
4 Complete understanding of the problem

Solving the problem

0 No attempt
1 Totally inappropriate plan
2 Partially correct procedure but with major fault
3 Substantially correct procedure with minor omission or procedural error
4 A plan that could lead to a correct solution with no arithmetic errors

Answering the problem

0 No answer or wrong answer based upon an inappropriate plan
1 Copying error; computational error; partial answer for problem with multiple answers; no answer statement; answer labeled incorrectly
2 Correct solution

problems is exacerbated by the failure of students to communicate clearly what they have done or what they are thinking. Students are prone to make calculations without explanations, and calculations alone often fail to reveal sufficiently the nature of the solver's work and thinking. It is not enough to check for right and wrong answers or to use multiple-choice formats for assessment of problem solving. As Silver and Kilpatrick (1988) state:

> A reliance solely on the sleek efficiency of multiple-choice (and other short answer) formats will severely hinder efforts to help students develop a reflective and interrogatory stance toward their learning.

If we can devise methods for eliciting better communication of students' thinking, we can perform more effective assessment. Such assessment measures the quality of students' thinking. This information can help teachers design and implement instruction to promote greater success in problem solving and can help administrators evaluate programs and curriculum.

Assessment of Solved Problems

The most natural and common method for assessing performance in problem solving is to obtain general impressions about the quality of a solution while scanning students' work. These general impressions are strongly influenced by the "proximity of correctness" of the answer. As a result, good solutions with minor errors due to carelessness that alter the answer dramatically can receive undeservedly low scores. Scales are available that focus more attention on solution procedures, enabling teachers to obtain fairer and more reliable scores. For example, Charles, Lester, and O'Daffer (1987) devised a scale that assigns separate scores to each of three stages in problem solving: understanding the problem, solving the problem, and answering the question. Figure 1 shows a modification of their scale, with increased emphasis given to the understanding and solving stages (Wilson 1991).

The Charles, Lester, and O'Daffer scale and its modified forms are easy to use. An advantage of such a scale is that a teacher may focus on only one of the stages. For example, a teacher who is emphasizing strategy selection and implementation can assess each student's solving procedure irrespective of the answer.

The California Assessment Program (Pandey 1990) includes comprehensive descriptions of various levels of performance for specific problems. This is appropriate for large-scale assessment programs. However, the classroom teacher has little time to construct scales for individual problems. Teachers need assessment procedures and scales that they can modify or use intact for a wide range of problems.

Categorizing Responses to Problems

Scales for assessment of problem solving can be designed without creating an evaluative threat to students. Such a system of scales was constructed for use in the 1990 British Columbia assessment of problem solving (Szetela 1991). Instead of scoring the solutions only, teachers analyze the responses to problems on the basis of four categories: answers, answer state-

Students are prone to make calculations without explanations, and calculations alone often fail to reveal sufficiently the nature of the solver's work and thinking.

FIGURE 2

CATEGORIES OF RESPONSES IN SOLUTIONS TO PROBLEMS

Answer
1. Blank
2. Undetermined
3. Incorrect
4. Correct

Statement
1. No statement
2. No context
3. No units
4. None required
5. Complete

Strategy Selected
1. Number sentence
2. Select operations and calculate
3. Algebraic
4. Non-systematic list
5. Systematic list
6. Guess and test
7. Draw diagram
8. Look for pattern
9. Logical reasoning
10. Use simpler case
11. Work backwards
12. Undetermined

Implementation
1. No work shown
2. Identifies data only
3. Problem misinterpreted
4. Strategy not clear
5. Strategy initiated (table, graph list) but incomplete or poorly implemented
6. Conditions or possibilities overlooked
7. Multiple secondary errors
8. A single secondary error
9 Appropriate and complete

ments, strategy selection, and strategy implementation (see fig. 2). Teachers can use a single category to determine how well their students are addressing a particular aspect of solving problems. One focus might be on strategies used. Another might be directed toward answer statements. Incomplete statements that fail to include the units taught or important contextual information may serve as focal points for teachers in their subsequent instructional activities.

Promoting Greater Communication

To further enhance assessment, we need to devise problem situations and questions that encourage and motivate students to communicate and explain their thinking. Figure 3 shows one way to do this. An already solved problem with a significant error, combined with a set of relevant questions about the solution, facilitates communication. As with an unsolved problem, students must form a suitable representation of the problem. Instead of solving the problem themselves, however, they analyze the given solution. Finally, they reveal their thinking by answering the pertinent questions. Answers to these questions can provide more comprehensive insights about the student's thinking in problem situations than more typical problem formats, in which students may have various levels of success but fail to reveal their thinking.

Assessment of responses to the questions accompanying the already solved problem can be done in less time than it normally takes for teachers to plod through the usual wide range of solution procedures for a given problem. The main goal of the example in Figure 3 is to determine whether or not a student understands a problem situation well enough to recognize the incongruity of the given answer despite excellent implementation of a good plan, with the problem solver running awry only in the careless writing of the answer statement. Teachers can provide continuing experiences for students to critically analyze solutions and communicate their observations and responses to relevant questions. Such practice can help students engage in reasoning, evaluating, and communicating, and can enable teachers to assess these problem-solving processes more effectively.

Other forms of problems with questions to stimulate thinking and written communication include the following:

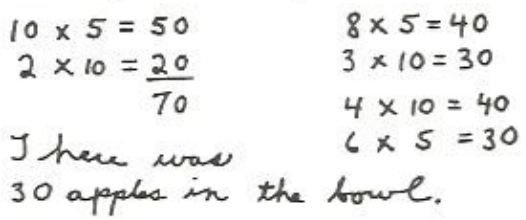
- Present a problem with all the facts and conditions, but have the student write an appropriate question, solve the completed problem, and write their perceptions about the adequacy of the problem.
- Present a problem and a partial solution. Have students complete the solution.
- Present a problem with facts unrelated to the question. Have students comment about the quality of the problem or revise the problem to remove the incongruity.
- Have students explain how they would solve a problem using only words, then solve the problem and construct a similar problem.
- After students solve a problem, have them write a new problem with a different context but preserving the original problem structure.
- Present a problem without numerals. Have students supply appropriate numerals, estimate answers, and solve the problem.

Teachers can assess the quality of

FIGURE 3

Example of Problem That Asks Students to Communicate Thinking

A bowl contains 10 pieces of fruit (apples and oranges).
Apples cost 5 cents each and oranges cost 10 cents each.
All together the fruit is worth 70 cents.
We want to find how many apples are in the bowl.
Kelly tried to solve the problem this way.



Try to follow Kelly's work and solution. Then answer the questions.

1. Is Kelly's way of solving the problem a good one?

   Yes

   Tell why you think it is or is not a good way.

   because it will tell you the posible silution which is what you want but she didn't read carfully

2. Did Kelly get the right answer?

   No

   Explain why she did or did not.

   because there are only ten items in the basket

each response by using a scale such as the following:

1. No response or simplistic or irrelevant response.
2. A relevant response but of minor importance with respect to the question or problem.
3. A reflective and significant response but with an important omission or misconception.
4. A comprehensive, logical, and correct response to the question or problem.

These suggestions for assessment of problem solving have the potential to reveal much more than we currently know about students' thinking, their conceptions, their weaknesses, and their strengths. With better awareness about students' knowledge and thinking, teachers can plan more effective instruction, and the outcome is more likely to be better learning of higher-order skills essential to success in problem solving.

References

Charles, R., F.K. Lester, Jr., and P. O'Daffer. (1987). *How to Evaluate Progress in Problem Solving.* Palo Alto, Calif., Dale Seymour Publications.

Garofalo, J., and F.K. Lester, Jr. (1985). "Metacognition, Cognitive Monitoring, and Mathematical Performance". *Journal for Research in Mathematics Education* 16: 163-176.

Pandey, T. (1990). "Power Items and the Alignment of Curriculum Assessment." In *Assessing Higher Order Thinking in Mathematics*, edited by G. Kulm. Washington, D.C.: American Association for the Advancement of Science.

Silver, E.A., and J. Kilpatrick. (1988). "Testing Mathematical Problem Solving." In *The Teaching and Assessing of Mathematical Problem Solving*, edited by R.I. Charles and E.A. Silver. Reston, Va.: The National Council of Teachers of Mathematics.

Szetela, W. (1991). "Open-Ended Problems." In *The 1990 British Columbia Assessment of Mathematics: Final Report,* edited by D. Robitaille, Victoria, B.C.: Ministry of Education Assessment Branch.

Wilson, D. (1991). "Analysis of Students' Problem Solving." Unpublished paper.

Walter Szetela is an Associate Professor and **Cynthia Nicol** is a mathematics teacher and a Ph.D. candidate in mathematics education. They can be reached at The University of British Columbia, Faculty of Education, Department of Mathematics and Science Education, 2125 Main Mall, Vancouver, B.C., CANADA V6T 1Z5.

Reprinted with permission from Educational Leadership (May 1992).

WOO YOKE YOONG

ASCD Diary of Professional Development

14-15 February 1992

Dimensions of Learning

Teaching of Thinking

About five hundred educators turned up at the NPB Auditorium to learn more about how children learn and think. Dr Ron Brandt, the Executive Editor of *Educational Leadership* and other ASCD (USA) publications gave a review of the various ideas in teaching thinking. He also introduced Singapore educators to current trends in Performance Assessment in America.

His co-speaker, Dr Debra Pickering gave the participants of the ASCD conference a preview of the Dimensions of Learning programme which would be published in USA in the second half of 1992.

1-3 July 1992

Cooperative Learning

As a result of the enthusiastic response ASCD members gave to Barbara Allen's introductory talk in 1991, ASCD Singapore asked Barbara Allen to conduct two workshops in 1992. The aim of these workshops was to give ASCD members hands-on experience in cooperative learning. "Ms Allen has tremendously increased my repertoire of skills in managing learning in groups", said a participant of the workshop.

Dr Ron Brandt

4 July 1992

Inviting School Success for Everyone

Professor William Purkey was another distinguished speaker who was brought back by popular demand. Over 650 teachers, heads of departments and administrators signed up to listen to the Professor speak on inviting pupils to realise their full potential. The Professor also shared his invitational approach to conflict resolution by describing what educators can do to solve discipline problems with the least amount of energy and in the most humane and dignifying manner. To do this, the Rule of the Five C's can be used. The five C's are,"Concern," "Confer," "Consult," " Confront," and "Combat."

10-11 Sept 1992

Be a Master Teacher

Singapore was privileged to hear the Professor Madeline Hunter, the guru of Mastery Teaching herself share her model on teaching.

"A typical teacher makes over 5000 professional decisions a day. Most of these are automated but they are not unconscious," revealed Prof Hunter. In the course of four lectures, Prof. Hunter showed how teachers could make more effective decisions by translating the findings of brain research, by knowing the conditions which promote or discourage transfer and by mastering the techniques

Professor William Purkey

Professor Madeline Hunter

of effective practice.

24-27 November 1992

Early Childhood Workshop Series

The Early Childhood Workshop Series are specially organised to cater to teachers and parents of young children. The next series of workshops are scheduled for the 24-27 November 1992. The panel of speakers for this series include

- Ms Betty Yeoh, Specialist Inspector (Science), Ministry of Education. Betty will talk about *Fun With Science*.
- Mrs Kamala Thiagarajan, Project Director of Primary English Programme, Curriculum Development Institute of Singapore. She will speak on *Shared Reading*.
- Ms Loh Geok Chin, formerly Vice Principal of Fuchun Primary. Ms Loh, who retired last year, will speak on *Fun With Numbers*.
- Dr Low Guat Tin, lecturer in the Division of Policy and Management Studies, National Institute of Education, will conduct a workshop on *Relating to Children*.

Woo Yoke Yoong is the Honorary Secretary of ASCD Singapore.

CAROL B FURTWENGLER

How to Observe Cooperative Learning Classrooms

Administrators who bring knowledge of various teaching styles to their classroom observations support teachers in expanding their repertoire of effective teaching practices.

Oh, it's time for the principal to conduct a formal observation of my classroom. Let's see - what lesson do I have that will follow those steps?

Such thoughts often permeate the thinking of teachers contemplating an administrator's imminent formal classroom observation. A prime reason for this is that a directed-teaching model was widely used during the 1980s to train administrators to conduct effective classroom observations; now evaluators look for the elements of effective instruction - a model that includes prescribed steps in each lesson (Juska 1991). It is time that administrators increase their observational skills beyond the directed-teaching model and incorporate other styles of teaching into their formative observational methods. Administrators need a repertoire of observational skills that will encourage teachers to use varying styles of teaching for improving instruction.

Cooperative learning is one teaching style that today's administrators must know. Research reveals that it improves students' academic achievement and social skills, and that is a popular style with students.

This article provides a framework for administrators to use when observing teachers who use cooperative learning in their classrooms. It presents a brief outline of popular models to acquaint administrators with their titles, developers, and major program characteristics (see fig. 1). The administrator can prepare for the observation using a set of suggested "discussion leads" for a preobservation conference. A set of explicit questions serves as a guide for an administrator to review before observation and to use in preparing feedback to teachers.

Cooperative Learning Programs

Numerous program designs for cooperative learning exist that can be used in various subject areas and different types of classrooms. The literature is replete with acronyms (STAD, TGT, AI, CIRC) and special learning methods (Jigsaw, Jigsaw II, Learning Together, Group Investigation). Administrators should not be discouraged by the multiplicity of designs for cooperative learning. Figure 1 outlines popular cooperative learning models by title (including acronyms), authors and a brief description of program characteristics. Two major purposes for cooperative learning programs - to improve student achievement and to increase social skills of students - are reflected in these programs. Robert Slavin and his colleagues (1984, 1986, 1990) at Johns Hopkins University develop programs that focus on cooperative learning that improves student achievement. David Johnson and Roger Johnson (1987, 1989, 1991), two brothers who share research interests, focus on techniques to improve students' social skills.

Preparing for Classroom Observation

Administrators who plan to observe a lesson in a cooperative learning classroom will benefit from a preconference with the teacher, which is an opportunity for the teacher and administrator to become "instructional colleagues." A preconference allows the teacher to share information about instruction and student learning and the administrator to gain additional knowledge about the classroom and identify a focus for the observation.

During the preconference, the administrator can learn the basic elements of the cooperative learning model the teacher is using and the point the students are at in their cooperative learning tasks. If, for instance, the teacher is using Cooperative Integrated Reading and Composition (CIRC), the administrator can expect to see the teacher direct some small-group instruction while

other pairs of student work independently on their assignments.

The administrator also discusses with the teacher the development of the cooperative learning lesson.

Suggested discussion leads include:

- Which cooperative learning model are you using? Why did you select this model?
- How did you form your groups?
- What objectives, time lines, and directions did the groups receive?
- What have you observed about group-processing roles of individual students? About academic achievements of individual students and groups?
- What task will the groups be engaged in during this observation?
- How can I best help you in this process?

The Observation of Cooperative Learning Instruction

The questions below provide a reference for determining the focus of the observation and for deciding what kind of feedback to give the teacher. Items are listed in the categories below:

Classroom Organization

Group size/composition. Does the group size match the cooperative learning model that the teacher is using, or is the size appropriate to the assigned task? Is group composition heterogeneous?

Room arrangement/materials. Are desks and chairs arranged so that all group members can see and hear one another? Is there adequate space for each group? Are materials for the lesson appropriate, available, and easily accessible? Do they promote cooperative learning activity?

FIGURE 1

PROGRAM CHARACTERISTICS OF POPULAR COOPERATIVE LEARNING MODELS

Student Teams Achievement Divisions (STAD) (Slavin 1986)
- Four-member, heterogeneous learning teams; designed for well-defined objectives.
- Direct instruction by teacher followed by work in student teams for mastery.
- Individual student quiz scores; then summed for team scores.

Teams-Games-Tournament (TGT) (DeVries and Slavin 1978)
- Like STAD but replaces quizzes with weekly, three-person "tournament tables"
- Teams matched against others of similar ability.
- Student teams regrouped each week based on individual performance.

Team Assisted Individualization (TAI) (Slavin et al. 1984)
- Four-member, heterogeneous teams for math, grades 3-6.
- Teacher instructs homogeneous students from all groups; students go back to teams to work.
- Team members work on individual units at their skill level but help each other.
- Individual unit tests taken without team help; weekly team awards.

Cooperative Integrated Reading and Composition (CIRC) (Stevens et al. 1987)
- Four-member, upper elementary teams; two members have same ability level.
- Teacher instructs pairs of similar ability (reading, writing and language arts).
- Team scores based on individual scores.

Jigsaw (Aronson et al. 1978)
- Six-member, heterogeneous teams, grades 3-6.
- Each team member learns assignment by becoming "expert" with members of other teams.
- Team members return to groups as "experts" and teach one another.

Jigsaw II (Slavin 1986)
- Four- to five-member teams.
- Students learn common material but become "expert" on subtopic; meet with "experts" on other teams; return to original team to teach material.
- Individual student quizzes with team results based on improvement.

Learning together (Johnson and Johnson 1987)
- Four- to five-member, heterogeneous groups, grades 2-6.
- Total class instruction by teacher; student groups work on assignments.
- One final product for team score.

Group Investigation (Sharan and Sharan 1980, 1989)
- Two- to six-member student groups.
- Groups choose topic and then assign individual tasks.
- Groups make presentations to entire class; receive group award.

Classroom Management

Clear guidelines established. Does the teacher establish clear guidelines that facilitate positive interdependence and promote group harmony? Is it clear everyone should contribute, help, listen with care to others, encourage others to participate, and ask for help or clarification?

Administrative procedures. Does the teacher establish and consistently enforce a set of rules and procedures that govern the handling of routine administrative procedures, student oral participation, and movement during different types of activities?

Transitions. Do smooth transitions occur, and do they culminate in students being ready to begin and finish work on their assigned task(s)?

Use of time. Does the teacher promptly start relevant administrative procedures such as roll call and begin instruction or provide directions for group work? Does the teacher keep students/groups actively involved in appropriate instructional tasks during the whole lesson?

Presentation of Content

Motivation. Does the teacher identify for students the importance and usefulness of the objective outlined at the beginning of the lesson? (This may occur in cooperative brainstorming, group discussions, or as part of instructional input offered by the teacher or students.) Do students discover what the topic is about, why it is interesting to them, and what they already know about the topic? Do students/groups demonstrate a high level of motivation and enthusiasm for the assigned task and in accomplishing group goals?

Input/modeling/review. Do the students (or teacher) provide input, when necessary, and encourage group members to use similar strategies? Are instructional examples provided by teacher, textbook, instructional media, and students? Do students discuss ideas in language familiar to their peers? Does the teacher explain relevant material and skills to the class, or do students offer explanations that relate lesson objectives to their knowledge and experiences? Do student groups use a variety of skills, such as reasoning, hypothesizing, predicting, and intuitive thinking? Do students check one another for understanding of concepts and skills and provide review, when necessary?

Group Facilitation

Cohesiveness. Do students show mutual respect for those of other races, ethnic origins, and social classes? Are students encouraged to work productively in their groups and reinforce (praise, reward) students who engage in appropriate behavior? Are students aware that they play a unique role on the team and that the team could not succeed without them? Do students "coat tail," or is each member of the learning team actively involved in the assigned task? Can students resolve conflicts constructively?

Clear role expectations. Are roles such as reader, recorder, calculator, checker, reporter, time-keeper, and materials handler or skill roles such as encourager of participation, praiser, and checker for understanding assigned during group work?

Accountability. Are students held accountable for individual learning through testing, individual work, or structuring activities so that each student is responsible for a specific part of the group product? Is the group accountable for its work and for the achievement of each member of the group? Does the cooperative learning experience focus the classroom reward system on helping others learn? Does the collaborative, rather than competitive mode, dominate?

Administrators need a repertoire of observational skills that will encourage teachers to use varying styles of teaching for improving instruction.

Monitoring

Intervening. Does the teacher monitor group progress and intervene when serious problems hamper group or individual learning? Does intervention, if necessary, assist groups in solving their problems, rather than "taking on the problem" for them?

Notes progress/problems. Does the teacher circulate, making note of individual/group accomplishments, how progress is being made toward goal attainment, and how problems are being resolved? Does the teacher provide task assistance by clarifying, reteaching, or elaborating?

Reteach/discussion. Does the teacher use notes from monitoring and student/group input to identify areas that need reteaching or further discussion? If problems or incorrect answers are discovered, does the teacher use this opportunity to reteach or discuss the correct answer or solution with the group? If problems occur in group interaction or work process, does the teacher review and reteach the social skills necessary to increase group cohesiveness and effectiveness?

Lesson Summary

Process/product effectiveness. At the conclusion of the group activity/project, do the students and teacher evaluate the progress made by the group (social and academic) and evaluate learnings (products/outcomes) from the student work?

Becoming Colleagues

Administrators who conference with teachers, understand the instructional model they are using, and determine how cooperative learning functions in the classrooms become teachers' instructional colleagues. By becoming familiar with the differences between a teacher-directed lesson and a cooperative learning lesson, they recognize teacher monitoring, teacher intervening, student group work, and student interaction as essential elements in the cooperative learning process, and they can provide genuine feedback to teachers about their classrooms. And teachers who recognize that administrators are interested in and knowledgeable about their instructional methods can use more diverse and more effective teaching practices during classroom observations.

References

Aronson, E., N. Blaney, C. Stephan, J. Sikes, and M. Snapp. (1978) *The Jigsaw Classroom.* Beverly Hills: Sage Publications, Inc.

DeVries, D.L., and R.E. Slavin. (1978). "Teams-Games-Tournament (TGT): Review of Ten Classroom Experiments." *Journal of Research and Development in Education* 12:28-38.

Johnson, D.W., and R.T. Johnson. (1987). *Learning Together and Alone.* 2nd edition. Englewood Cliffs, N.J.: Prentice-Hall.

Johnson, D.W., and R.T. Johnson. (1989). *Cooperation and Competition: Theory and Research.* Edina, Minn.: Interaction Book Company.

Johnson, D.W., R.T. Johnson, and E.J. Holubec. (1991). *Cooperation in the Classroom.* rev. ed. Edina, Minn.: Interaction Book Company.

Juska, J. (1991). "Observations." *Phi Delta Kappan* 72: 468-470.

Sharan, S. (1980). "Cooperative Learning in Small Groups: Recent Methods and Effects on Achievement, Attitudes, and Ethnic Relations." *Review of Educational Research* 50, 2: 241-271.

Sharan, Y., and S. Sharan. (1989). "Group Investigation Expands Cooperative Learning." *Educational Leadership* 47: 17-21.

Slavin, R.E. (1986). *Using Student Team Learning.* 3rd ed. Baltimore: Johns Hopkins University, Center for Research on Elementary and Middle Schools.

Slavin, R.E. (1990). *Cooperative Learning: Theory, Research and Practice.* Englewood Cliffs, N.J.: Prentice Hall.

Slavin, R.E., N.A. Madden, and M. Leavey. (1984). "Effects of Team Assisted Individualization on the Mathematics Achievement of Academically Handicapped Students and Nonhandicapped Students." *Journal of Educational Psychology* 76: 813-819.

Stevens, R.J., N.A. Madden, R.E. Slavin, and A.M. Farnish. (1989). "Cooperative Integrated Reading and Composition: Two Field Experiments." *Reading Research Quarterly* 22: 433-454.

Author's note: Graduate students in clinical supervision classes at Wichita State University worked in cooperative groups on this project and contributed to the content of the article.

Carol B Furtwengler is Associate Professor, Wichita State University, 115 Hubbard Hall, Box 142, Wichita, KS 67208.

Reprinted with permission from Educational Leadership (April 1992).

CAROLYN M EVERTSON AND ALENE H HARRIS

What We Know About Managing Classrooms

Effective classroom management must move beyond the control of behaviors. Future research needs to describe how to create supportive learning environments in schools that face complex and changing needs.

For the last 15 years Gallup polls have reported the public's belief that the answer to many school problems is improved discipline. Among practitioners, particularly new teachers, classroom management and discipline remain their number one concerns (Veenman 1984).

Not surprisingly, the response in education research has been to expand our knowledge of what effective classroom managers do and how they do it. As a result of a broadened definition of classroom management, today's research moves away from a focus on controlling students' behavior and looks instead at teacher actions to create, implement, and maintain a classroom environment that supports learning (Johnson and Brooks 1979; see Brophy 1983, and Doyle 1986, for comprehensive reviews).

Our purpose is to review the evolution in research on classroom management, to investigate how this knowledge translates into real experiences for teachers, to suggest directions for further exploration.

Reviews of Research

Studies about time. "The association of learning with time is among the most consistent that education research reveals" (Walberg 1988, p. 84). Past research (Karweit 1988) indicates that:

- The amount of time students spend learning the curriculum varies from school to school.
- Even under the best of circumstances, half or less of the school day is used for instruction.
- The amount of instructional time spent is often associated with student achievement.

Although policymakers use these findings to support extending the school day, Walberg (1988) suggests that increases in productive time must accompany increases in allocated time.

While the amount of time available imposes outer limits on what can be accomplished, the key issue is really *how* time is used. Effective classroom management conserves instruction time by planning activities and tasks to fit the learning materials; by setting and conveying both procedural and academic expectations (constructing and teaching lessons on "going-to-school skills"); and by ap-

While the amount of time available imposes outer limits on what can be accomplished, the key issue is really how time is used.

propriately sequencing, pacing, monitoring, and providing feedback for student work (Emmer, Evertson and Emmer 1982).

Research indicates however, that teachers must be aware of and make visible what students are actually learning because students may seem involved in tasks without engaging in the content. Bloome, Puro, and Theodorou (1989) refer to this as "procedural display" and "mock participation" when students and teachers engage in activities without being involved in the content substance.

Group management strategies. Kounin (1970), reaffirmed by Gump (1982), identified several strategies that teachers use to elicit high levels of work involvement and low levels of misbehavior:

- *Withitness* - communicating awareness of student behavior;
- *Overlapping* - doing more than one thing at once;
- *Smoothness and momentum* - moving in and out of activities smoothly with appropriately paced and sequenced instruction; and
- *Group alerting* - keeping all students attentive in a whole-group focus.

Lessons that engage students. Certain class activities also elicit varying degrees of student engagement (Gump 1982). For example, Kounin (1970) found highest student engagement (85 percent) during recitation and lowest (65 percent) during seatwork. Other studies emphasize these findings, adding that:

- frequent seatwork results in lower on-task behavior (Anderson 1984);
- alternating cycles of two shorter segments each of content development and seatwork maintain higher student involvement than single longer sequences (Evertson 1982); and
- student engaged rates during seatwork differ among teachers, while engaged rates during recitation are similar (Edenhart-Pepe, Hudgins, and Miller 1981).

Recitation remains the dominant instruction method probably because it is an easier way to keep students involved, looks orderly, and seems equivalent to learning. Also, transitions required in more complex formats can result in lost time (Arlin 1979).

Teachers must recognize both academic and social dimensions of classroom tasks. For example, students have to interpret not only what they are to learn, but how they are to participate. Teachers need to make clear their expectations and procedures for student participation, for example, how to answer questions or bid for a turn (Green and Smith 1983;

Highlights or Research on Classroom Management

Research on classroom management must change focus to meet the complex needs of creating and maintaining an environment that supports all aspects of learning.
A review of past research, reinforced by field studies, confirms certain elements as basic to effective classroom management. Teachers who are effective managers:

- use time as effectively as possible;
- implement group strategies with high levels of involvement and low levels of misbehavior;
- choose lesson formats and academic tasks conducive to high student engagement;
- communicate clearly rules of participation;
- prevent problems by implementing a system at the beginning of the school year.

Future research needs to:

- define the impact of the school and community cultures on teacher efforts;
- ensure that the substance of what is being taught is adequate;
- define effective management techniques to fit the need for classroom management that encourages more problem solving and less routinized academic tasks.

More intellectually demanding academic work and activities in which students create products or encounter novel problems require more complex management.

Weade and Evertson 1988).

Assignments with varying cognitive and procedural complexity have consequences for classroom management. Generally, more intellectually demanding academic work and activities in which students create products or encounter novel problems require complex management decisions (Bossert 1979), which demonstrates the interrelated nature of classroom management and the curriculum (Doyle and Carter 1984).

Classroom communication. Studies about classroom communication - verbal and nonverbal ways that norms, rules, and expectations are signaled - show how both students and teachers actively mediate and construct the learning environment. See Erickson (1986) and Green and Smith (1983) for reviews. Studies identify what students need to understand and to participate in lessons and how teachers orchestrate that participation. Some ritualistic activities, such as passing out papers, require little understanding. However, nonritualistic activities require students to "read" the requirements correctly or risk negative evaluations of their behavior and abilities (Green and Harker 1982). A close look at how class activities evolve reveals the need for a classroom management system that is visible, established, monitored, modified, refined, and reestablished.

Teachers' managerial decisions. Expert teachers are influenced by a rich store of information that allows them to judge what are typical and nontypical classroom scenes (Carter 1990). They see classrooms as "moving systems" and make managerial decisions based on their perceptions of how well students are working within those systems. These teachers interpret and act on cues from students that signal students' involvement - or lack of it - in academic tasks (Carter, Cushing, Sabers, Stein, and Berliner 1988).

Beginning the year. Koumin (1970) suggested that it is not so much what teachers do to stop misbehavior that characterizes effective group management, but how they prevent problems in the first place. In response, several studies investigated how effective managers began their school year, and discovered that in both elementary and secondary classrooms the start of the school is crucial to effective management (Emmer et al. 1980; Evertson and Emmer 1982). Teachers whose students demonstrated high on-task rates and academic achievement implemented a systematic approach toward classroom management at the beginning of the school year. They began the year by:

- preparing and planning classroom rules and procedures in advance;
- communicating their expectations clearly;
- establishing routines and procedures, and teaching them along with expectations for appropriate performance;
- systematically monitoring student academic work and behavior; and
- providing feedback about academic performance and behavior.

In classrooms with this sort of system, there are improved student task engagement, less inappropriate behavior, smoother transitions between activities, and generally higher academic performance (Emmer et al. 1980; Evertson and Emmer 1982).

Handling misbehavior. Of course, a carefully planned management system will not, by itself, stop all misbehavior, but teachers can usually handle it unobtrusively with techniques such as physical proximity or eye contact. More serious misbehavior may require more direct intervention. Because punishment neither

The success of intervention depends on orderly structures being in place.

Research findings can and do influence teacher decisions about management practices.

teaches desirable behavior nor instills a desire to behave, it is perhaps best used as part of a planned response to repeated misbehavior. However, the success of intervention depends on orderly structures being in place. Well-understood norms and expectations for behavior must have previously existed (Doyle 1990).

The same hold true for discipline programs such as Teacher Effectiveness Training, Reality Therapy, and Assertive Discipline. These systems provide methods for dealing with threats to classroom order, but Emmer and Ausikker (1990) found that none adequately addresses the complex preventive and supportive functions necessary for effective management and discipline. They contend that these systems "[fail] to address the day-to-day classroom management skills needed to engage students in productive activities and to prevent minor problems from becoming major ones" (p. 146). Their analysis of 36 studies supports the need to establish a comprehensive system of management and organization early in the year.

The continuing exploration of behavior modification techniques, especially in special education, is shifting from teacher control to student self-monitoring and self-control. Teachers who apply Meichenbaum's (1977) ideas by using a combination of modeling and self-verbalization help aggressive students control anger, deal with frustration, and respond to errors with problem-solving efforts. In addition, as Doug and Lynn Fuchs and their colleagues at Peabody College, Vanderbilt, have found, a combination of goal-setting and self-recording techniques help at-risk students improve their behavior and academics and reduce special service referrals (Fuchs et al. 1990).

Educating Teachers in Classroom Management

While early studies provided generic information about classroom characteristics and teacher actions that produce order and student involvement, conceptual frames were needed to help teachers orchestrate these principles in the fast pace of a classroom.

Subsequent field studies supported the concept of establishing a classroom management system at the beginning of the school year. Teachers in the experimental groups not only used significantly more management strategies and procedures than comparison groups, but also their students exhibited higher task engagement, less inappropriate behavior, and higher academic success (Evertson, Emmer, Sanford, and Clements 1983; Emmer, Sanford, Clements, and Martin 1983; Evertson 1985; 1989).

These findings provide evidence that research findings can and do influence teacher decisions about management practices (Evertson et al., 1983; Evertson 1985, 1989; Putnam and Barnes 1984).

However, this is not the whole story. Learning to teach is a complex enterprise that requires practice in problem-solving more than acquisition of rote skills (Brophy 1988; Evertson 1987). To achieve that end, educators at Peabody College, Vanderbilt, are using videodisc technology to design problem-solving contexts requiring managerial decisions. See also Richardson (1990) for work on ways to use research as a base for teacher preparation.

Future Inquiry

As recent research indicates, three topics should be central to the future study of classroom management and discipline: school level discipline and classroom management; quality of academic tasks; and classroom management in different contexts.

School discipline and classroom management. Although not included in this review, abundant literature documents how school-level discipline influences classroom management (Moles 1990). The view

Good management and organization must focus more on the content and substance of what is being managed.

presented is that the school and community cultures affect the values and decisions teachers make in their management systems. In other words, teachers' management decisions that are not supported at the school and community levels lose credibility with students.

Quality of academic tasks. Good management and organization must focus more on the content and substance of what is being managed and less on the look of engagement. Whereas good classroom management is necessary for learning, it does not stand alone. Recent research has identified classrooms with high levels of student engagement, but meager academic content, resulting in low levels of learning (Weade and Evertson 1988).

Classroom management in different contexts. The current climate of school reform clearly calls for teaching problem-solving and higher order thinking skills, integrating learning experiences within and across subject areas, and implementing multiple tasks (Resnick 1987). Enacting these changes requires new methods of organization and management (Cohen and Lotan 1990; Marshall in press). Most classroom management studies have looked at classrooms with routinized, predictable academic tasks and activities. Little research has examined different instructional contexts, for example, whole language settings or process writing, and the managerial decisions required (Edelsky, Draper and Smith 1983).

All in all, future research needs to address these questions: How can classroom management and organization support students' substantive learning? And what is the nature and quality of the learning that is supported?

References

Anderson, L.W. (1984). *Time and School Learning: Theory, Research, and Practice.* London: Croom Helm.

Arlin, M. (1979). "Teacher Transitions Can Disrupt Time Flow in Classrooms." *American Educational Research Journal* 16, 1:42-56.

Bloome, D., P. Puro, and E. Theodorou. (1989). "Procedural Display and Classroom Lessons." *Curriculum Inquiry* 19, 3: 265-291.

Bossert, S.T. (1979). *Tasks and Social Relationships in Classrooms.* Cambridge: Cambridge University Press.

Brophy, J.E. (1983). "Classroom Organization and Management." *The Elementary School Journal* 83, 4: 265-285.

Brophy, J.E. (1988). "Educating Teachers About Managing Classrooms and Students." *Teaching and Teacher Education* 4, 1: 1-18.

Carter, K. (1990). "Teachers' Knowledge and Learning to Teach". In *Handbook of Research on Teacher Education,* edited by R. Houston. New York: Macmillan.

Carter, K., K. Cushing, D. Sabers, P. Stein, and D. Berliner. (1988). "Expert-Novice Differences in Perceiving and Processing Visual Classroom Information." *Journal of Teacher Education* 39, 3: 25-31.

Cohen, E.G. and R.A. Lotan. (1990). "Teacher as Supervisor of Complex Technology." *Theory into Practice* 29, 2:78-84.

Doyle, W. (1986). "Classroom Organization and Management." In *Handbook of Research on Teaching,* (3rd ed.), edited by M. Wittrock. New York: Macmillan.

Doyle, W. (1990). "Classroom Management Techniques." In *Student Discipline Strategies,* edited by O.C. Moles, Albany, N.Y.: State University of New York Press.

Doyle, W. and K. Carter. (1984). "Academic Tasks in Classrooms." *Curriculum Inquiry* 14, 2: 129-149.

Edelsky, C., K. Draper, and K. Smith. (1983). "Hookin' 'Em at the

Start of School in a 'Whole Language' Classroom." *Anthropology and Education Quarterly* 14, 4: 257-281.

Edenhart-Pepe, M.M., B.B. Hudgins, and D.M. Miller (April 1981). "Who is Engaged More - Teacher or Students? An Analysis of How Activity Structures Affect Student Learning Engagement." Paper presented at the annual meeting of the American Educational Research Association, Los Angeles.

Emmer, E.T., and A. Aussiker. (1990). "School and Classroom Discipline Programs: How Well Do They Work?" In *Student Discipline Strategies,* edited by O.C. Moles, Albany, N.Y.: State University of New York Press.

Emmer, E.T., C.M. Evertson, and L.M. Anderson. (1980). "Effective Classroom Management at the Beginning of the School Year." *The Elementary School Journal* 80, 5: 219-231.

Emmer, E.T., J.P. Sanford, B.S. Clements, and J. Martin. (March 1983). "Improving Junior High Classroom Management." Paper presented at the annual meeting of the American Educational Research Association, Montreal. (ERIC Document Reproduction Service #ED 234 021).

Erickson, F. (1986). "Qualitative Methods in Research on Teaching.: In *Handbook of Research on Teaching* (3rd ed.), edited by M. Wittrock, New York: Macmillan.

Evertson, C.M. (1982). "Differences in Instructional Activities in Higher- and Lower-Achieving Junior High English and Math Classes." *The Elementary School Journal* 82, 4: 329-350.

Evertson, C.M. (1985). "Training Teachers in Classroom Management: An Experimental Study in Secondary School Classrooms." *Journal of Educational Research* 79, 1: 51-57.

Evertson, C.M. (1987). "Creating Conditions for Learning From Research to Practice." *Theory into Practice* 26, 2: 44-50.

Evertson, C.M. (1989). "Improving Classroom Management: A School-Based Program for Beginning the Year." *Journal of Educational Research* 83,2: 82-90.

Evertson, C.M., E.T. Emmer, J.P. Sanford, and B.S. Clements. (1983). "Improving Classroom Management: An Experiment in Elementary Schools." *The Elementary School Journal* 84,2: 173-188.

Evertson, C.M. and E.T. Emmer. (1982). "Effective Management at the Beginning of the School Year in Junior High Classes." *Journal of Educational Psychology* 74,4: 485-498.

Fuchs, D., L. Fuchs, S. Gilman, P. Reeder, M. Bahr, P. Fernstrom, and H. Roberts. (1990). "Preferral Intervention Through Teacher Consultation: Mainstream Assistance Teams." *Academic Therapy* 25, 3: 263-276.

Green, J.L., and D. Smith (1983). "Teaching and Learning: A Linguistic Perspective." *The Elementary School Journal* 83, 4: 353-391.

Green, J.L., and J. Harker (1982). "Gaining Access to Learning: Conversational, Social, and Cognitive Demands of Group Participation." In *Communicating in the Classroom*, edited by L.C. Wilkinson. New York: Academic Press.

Gump, P.V. (1982). "School Settings and Their Keeping." In *Helping Teachers Manage Classrooms*, edited by D. Duke. Alexandria, Va.: Association for Supervision and Curriculum Development.

Johnson, M. and H. Brooks. (1979). "Conceptualizing Classroom Management." In *Classroom Management* (78th Yearbook of the National Society for the Study of Education, Part II), edited by D. Duke. Chicago: University of Chicago Press.

Karweit, N. (1988, February). "Time-on-task: The Second Time Around." *NASSP Bulletin* 72, 505: 31-39.

Kounin, J.S. (1970). *Discipline and Group Management in Classrooms.* New York: Holt, Rinehart and Winston.

Marshall, H.H. (Ed.) (in press). *Supporting Student Learning: Roots of Educational Change.* Norwood, N.J.: Ablex Publishing Corp.

Meichenbaum, D. (1977). *Cognitive Behavior Modification.* New York: Plenum.

Moles, O.C. (1990). *Student Discipline Strategies: Research and Practice.* Albany, N.Y.: State University of New York Press.

Putnam, J. and H. Barnes. (1984). "Applications of Classroom Management Research Findings." Research Series 154, Institute for Research on Teaching, East Lansing, Mich.: Michigan State University (ERIC Document Reproduction Service #ED 256 752).

Resnick, L.B. (1987). *Education and Learning to Think.* Washington, D.C.: National Academy Press.

Richardson, V. (1990). "Significant and Worthwhile Change in Teaching Practice." *Educational Researcher* 19, 7: 10-18.

Veenman, S. (1984). "Perceived Needs of Beginning Teachers." *Review of Educational Research* 54, 2: 143-178.

Walberg, H. (1988). "Synthesis of Research on Time and Learning." *Educational Leadership* 45, 6: 76-85.

Weade, R. and C.M. Evertson (1988). "The Construction of Lessons in Effective and Less Effective Classrooms." *Teaching and Teacher Education* 4.4: 189-213.

Carolyn M. Evertson is Professor of Education. **Alene H. Harris** is Research Assistant Professor of Education. They can be contacted at Box 541, Peabody College, Vanderbilt University, Nashville, TN 32703.

Reprinted with permission from Educational Leadership (April 1992).

LOOI CHEE KIT

General Trends in Intelligent Tutoring Systems

In this article, I will review the broad spectrum of research concerned with how artificial intelligence and other advanced technologies can be applied to education and training. I will discuss the current state of the art, some future research directions as well as the transition of Intelligent Tutoring Systems from laboratories to the schools and training centres.

Computer-Assisted Instruction and Intelligent Tutoring Systems

Computer-Assisted Instruction (CAI) has been with us for a good while now. It is based on the behaviourist tradition which was dominant in the 50s and 60s. A meta-study of thousands of CAI systems done by Jim Kulik of the University of Michigan has shown that on the average CAI can provide 30% speed-up in learning time and a 10% improvement on scores. This means CAI despite its various limitations can be appropriate (Soloway, 1992).

Research into Intelligent Tutoring Systems (ITS) started in the early 70s. Its goal is to individualize one-on-one tutoring. A study by Bloom (1984) shows that one-on-one tutoring is the most effective way of instruction compared with classroom instruction. The architecture of an ITS would typically consist of:

- a domain expert system, usually a glass box which can solve problems on its own and provide a trace of its problem-solving;
- a tutoring expert system;
- a model of the student;
- an interface module.

Technology such as Artificial Intelligence (AI) is applied to build such modules (see also Looi, 1991).

The underlying educational philosophy in CAI and ITS is knowledge transfer. The main differences between CAI and ITS are:

- The student model in the ITS is more fine-grained than that in CAI;
- An ITS has an explicit model of domain knowledge;
- An ITS has an explicit model of tutoring.

The result is that there is more flexibility in the ITS system's response. ITS are built based on information processing psychology (a la Newell & Simon) instead of the behaviorist psychology behind CAI (Soloway, 1992).

Over the years, many intelligent tutoring prototypes and working systems have been built. Some have been field-tested (e.g. in the US Air Force). In general, the findings are that students using ITS achieved 1-sigma improvements in scores as well as reduction in the time spent on doing the task.

A current trend in ITS research is to work towards how to make possible the engineering of ITS. The vision is

The current model of learning - knowledge transfer - which was appropriate in the industrial age, is no longer relevant in the information age.

one in which companies build ITS components or tools for building such components (that a teacher or instructor can use to create courseware).

Situated Learning

New ways of looking at education have recently emerged. This has implications on the way we design teaching programs. A recent attitude is that the current model of learning - knowledge transfer - which was appropriate in the industrial age, is no longer relevant in the information age. This accumulation model of learning assumes that the core of learning is the transfer of information from the teacher to the student. What has been the result? In the US, statements have been made by educators and politicians that kids there coming out of school are not just prepared for the workplace. The new model of education advocates situated learning which posits learning an active, constructive and social process, and the main mechanism for learning should be to learn by doing. Students should learn collaboratively instead of individually. In this respect, the role of technology is to enable and facilitate such learning to take place.

This new approach has forced the ITS community to reexamine some fundamental assumptions in ITS work and to work towards new paradigms and possibilities for using technology for learning. To quote from William Clancey (1992):

"Certainly ITS researchers are aware of the importance of "people-oriented and organizational issues." But doing this means much more than "treating subject matter experts as active team members." For an instructional program ... we must involve students, teachers, administrators, future employers, and the community. We must observe the program in everyday use. This requires a major leap from the experimental paradigm of testing ready-made programs on a few subjects in a computer laboratory: We must modify the serial process of design, implementation, and evaluation. As ITS matures, we necessarily broaden our goals from developing new representational methods, with emphasis on problem solving in individual cases, to changing practice - changing how people interact, and changing their lives. We must move from evaluating isolated cognitive capabilities of the program (e.g. how well the explanation works) to designing new socio-technical systems. Design becomes an integrative process with many competing voices. How to design a system that people actually use becomes our research focus."

New developments in learning research have suggested that knowledge is not just in the head, but it also arises through interaction and negotiation with the community. For example, an examination of everyday activities shows how people ease the burden of cognition in particularly useful ways that are not recognized by standard teaching or AI methodology (Brown, 1990). We all use our embedded position in the world to off-load onto our environment part of the representational and the computational burden of cognition. The processes that we use to do this allow us to respond in real time to events as they happen in the world in which we are in. Consider this example (from Brown, 1990) which describes a study of computation in a group of people who were preparing carefully regulated meals:

"In this case they were to fix a serving of cottage cheese, supposing the amount laid out for the meal was three-quarters of the two-thirds cup the program allowed. The problem solver in the example began the task muttering that he had taken a calculus course in college ... Then after a pause he suddenly announced that he had "got it!" From then on he appeared certain he was correct, even before carrying out the procedure. He filled a measuring-cup two-thirds full of cottage cheese, dumped it out on the cutting board, patted it into a circle, marked a cross on it, scooped away one quadrant, and served the rest.

We note that the dieter's solution path was extremely practical. It reflected the nature of the activity, the resources available, and the sort of resolution required. The dieter's position gave him the ability to exploit the particulars of the context in which he was embedded and a privileged access to the solution path he chose. He was able to see the problem and its resolution in terms of the measuring cup, cutting board, and knife. Apparently inconsequential parts of the environment were appropriated as computing tools. He thus could use his environment to share with him with the representational and computational burden.

Situated learning raises a whole range of issues on education. For example with regards to teaching programming, is doing a skill like programming as a student in the school or university like doing programming professionally? Do we teach them enabling skills or should the student be given the opportunity to act as a real programmer?

On the average, Computer-Aided Instruction can provide 30% speed-up in learning time and a 10% improvement on scores.

Interactive Learning Environments

Systems embodying this situated learning approach include Interactive Learning Environments (ILEs). ILEs facilitate learning through the building of artifacts, facilitate both solo and collaborative work, and provide context by using video and audio media. ILEs need not embody any expert system - they could provide:

- an environment whereby one student tries to solve a problem on a computer and another student critiques or collaboratively helps to solve the problem on another linked computer;
- an environment whereby the computer learns collaboratively with the student;
- an environment that facilitates group activities;
- an environment for apprenticeship learning where the system solves or poses for the student the most difficult problems that occur;
- and other possibilities.

To contrast, the educational goal in CAI and ITS is the accumulation of knowledge through instruction, while the model of ILEs emphasizes doing the process through construction activities. In this article, we will continue to use ITS as a general term to mean and cover the broad spectrum of systems that include ITS and ILE.

A current viewpoint is that we need all of these different paradigms for learning. This is highlighted in the work done by Alan Lesgold and his group at the University of Pittsburgh who developed a few systems for use in the US Air Force. The systems follow a learning-by-doing approach. However, Alan Lesgold noted that learning by doing is not enough since the student does not always learn by doing discovery by himself. Noting that students also learn by collaboration and peer critiques, he called for incorporating different perspectives for learning such as learning by reacting and learning by critique. To do this, we need to incorporate multiple experts and multiple models in simulation, troubleshooting, coaching, modelling and monitoring.

The Experiences of Others *(or If ITS are such good ideas, why aren't there more of them?)*

Over the past eight years, John Anderson and his group at Carnegie Mellon University have been working on three computer-based tutors - one for beginning coding skills in the computer programming language LISP, one for proof skills in geometry, and one for symbol manipulation skills in algebra (Anderson, 1992). The tutors were completed by 1987. The major finding, which has been replicated many times, is that these tutors could accelerate the rate of learning by as much as a factor of three. In spite of this, the geometry and algebra tutors were a failure after they were both demonstrated in the Pittsburgh Public Schools over a period of three years. The perceived reasons were for the failure of the geometry tutor are:

- The way geometry was taught in the tutor is different from the way geometry was taught in the schools;
- There were no suitable machines for running the tutor;
- The tutor did not fit into the geometry curriculum and classroom;
- There was a lack of promotional effort.

In the evaluation of the algebra tutor, the experimental group was shown to have no advantage over the control group. The algebra tutor failed because:

- as the teachers were not involved in the design of the algebra tutor, they resisted the use of the tutor;
- the interface was too complex for the students.

Learning from these experiences, John Anderson and his group are now working with the Pittsburgh Public Schools to help revise the high school mathematics curriculum to one that is more modern, computer-intensive, and organized around computer tutors. They are now working with the mathematics faculty of one high school to develop other tutors. John Anderson noted that this is largely a matter of good intention now.

NYNEX Science & Technology, Inc. has developed an ITS for teaching COBOL (called Grace) to both novice and experienced programmers (McKendree et al, 1992). The goal was to take the current state of the art technology and apply it to the development of a tutor which can be used in real classrooms and in a variety of settings. The Grace project has been a success in the sense that its developers have been able to take technology, adapt it to their needs, and field an effective tutor in several classes. However, it is still not an unqualified success in that it has not been used outside of the lab setting and trials. The developers of Grace noted that they have still not clear the major hurdle of having the tutor taken from their hand and used every day in classrooms. They asked: How can a state-of-the-art system, with demonstrated success, not be readily accepted into the classroom? Four important reasons have been suggested: the lack of resources, equipment, personnel, and proper system developer attitudes. However, Grace's developers truly felt that they satisfied these criteria and that the real issue is the lack of effective transformation of the organization and the people involved. Neither they nor their end users put forth enough effort to make sure that the organiza-

tion was well prepared to take the final step. The problem as they see it is one of introducing new technologies into old environments. ITS are able to increase students' scores, reduce training time, and help the student migrate from the classroom to the job sites by providing situated learning opportunities. They are not simply incremental improvements of an existing process; they create a new way to deliver instruction.

Grace's developers also thought that their failure to get ITS into the classrooms is also largely a failure of their approach to technology transfer. The accumulation model of technology transfer works no better than the accumulation model of learning: practitioners are not empty vessels into which researchers funnel facts or technology. They felt that they must involve the practitioners as active and valued participants so that they may learn by doing as well.

Research Directions

The movement on situated learning is growing in the ITS community. However, there is still a large group (perhaps the predominant group) in the ITS community who would not agree with the shifting of the emphasis in ITS research. They still believe and work on mainstream ITS ideas - domain models, student modelling and tutoring which are indeed difficult issues. One way to put these differing approaches in perspective is to consider mainstream ITS work as basic research in the sense that we still do not know enough about learning; and to regard working with the community of users advocated in situated learning as applied research in the sense of applying successful mainstream ITS results.

While the situated learning perspective calls for working with the user community to facilitate learning using technology, the mainstream ITS community feels that this is just not the full picture. To them, it depends on what you really want to do. One could still work in a research laboratory and have minimal contacts with the real world if one is researching into (say) educational theory. Or one is a computer scientist or AI person who works on abstract modelling and believes that the work, while theoretical, is relevant just as one need mathematical models if we want to do an engineering project.

A speaker in the Second International Conference on Intelligent Tutoring Systems held in Montreal in 1992 brought out an important message: who gets to define what progress in this field is? His answer is: it depends. To students, does it make them learn faster and better? To teachers, progress is what makes their teaching easier. To computer scientists, is the code fast and compact? To the psychologists, is the stuff plausible? To testing agencies, does it save money? It is important to bear this perspective in mind as we approach any ITS work.

When ITS research started, many promises of building effective tutoring systems were made, and after some twenty years, there are still not many showcases of successful systems. This may explain a rush to build prototypes quickly and thereby shortchange long-term fundamental research goals.

ITS are considered as the more sophisticated and complex branch of knowledge-based systems (KBS). The practical impact of ITS in educational automation is in no proportion to that of KBS in other areas. This may be due to the complexity of ITS and to underdeveloped technology.

Applied ITS engineering does not exist yet, but there are at least some initial steps, e.g. in the work of Anderson's group, Woolf's group and commercial spin-offs of ITS research groups. The views and techniques that come out of ITS research do not constitute by themselves a comprehensive technology or methodology for building ITS. In the case of the EUROHELP[1], the engineering perspective drove the researchers deeper and deeper into fundamental research. This would seem to suggest that AI is not ready yet to support full

When ITS research started, many promises of building effective tutoring systems were made, but after some twenty years, there are still not many showcases of successful systems.

scale KBS or ITS applied engineering. This is only partially true, because one may easily point to areas where AI techniques could be effectively exploited, but where application is incidental.

For ITS to go practical, it should develop well-founded methodologies. The grounding of these methodologies leads inevitably to fundamental questions and research. The results of engineering-oriented projects and research into methodologies for knowledge engineering may be major sources of inspiration.

One such methodology may be provided by case-based reasoning technology (CBR). CBR is a model of cognition and learning that suggests the goal of a case-based ITS should be to teach cases as well as how to index them. In this sense, CBR is a technology for building one flavour of ITS. It stresses the construction, indexing, and use of large libraries of related cases, rather than the development of domain expert systems (Riesbeck & Schank, 1991).

ITS research grew as a testing ground of AI research into a semi-independent field (AI & Education) and contributed in this way to AI. Educational needs in society lead to a demand for (cost-)effective automation. Large sums of money are spent yearly by corporations in the United States on education & training:

IBM	$2 billion
AT&T	$700 million
military	$20 billion
Xerox	$10 million
Bellcore	$1 billion

This poses a challenge to use technology to improve training and thereby reduce costs.

The Future

ITS is a field "caught between science and engineering." Many aspects of ITS are still fundamental research problems. From a technological perspective, some technology is already there and ready for use, and technologists can work with users and practitioners on ways to exploit these technology for facilitating learning. From a applied research perspective, the technologist's role would be to take successful research results from the ITS community and think of ways we can apply them.

Multimedia is a facilitator for presenting different forms of media information. It does not offer any suitable viewpoint on how to approach a learning activity except for a disposition towards exploratory learning. In this respect, ITS and ILE have much to offer in adopting viewpoints on how learning should proceed with the assistance of multimedia presentation.

In conclusion, the question we want to ask ourselves is: how do we use the technology to effect a change in the learning culture in our schools, in the home, at the workplace, and in public places, to promote more effective and efficacious learning. The task is enormous and a challenging one.

[1] An ESPRIT funded project that involved more than a hundred person-years on research into a shell for building intelligent help systems for users of conventional software applications.

References

J.R. Anderson (1992), "Intelligent Tutoring and High School Mathematics", Lecture Notes in Computer Science 608, Springer-Verlag.

B.S. Bloom, (1984), The 2-Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-one Tutoring, Educational Researcher, Vol. 13, No. 13.

J. Breuker (1992), "Generality Watching: ITS Caught Between Science and Engineering", Lecture Notes in Computer Science 608, Springer-Verlag.

J.S. Brown (1990), "Toward a New Epistemology for Learning", ITS: At the Crossroads of AI and Education, Ablex Publishing.

W.J. Clancey (1992), "Guidon-Manage Revisited: A Socio-Technical Systems Approach", Lecture Notes in Computer Science 608, Springer-Verlag.

C.K. Looi (1991), "Artificial Intelligence Meets Education:An Overview", Association for Supervision and Curriculum Development (Singapore) Review, July 1991 issue.

J. Mckendree, B. Radlinski & M.E. Atwood (1992), "The Grace Tutor: A Qualified Success", Lecture Notes in Computer Science 608, Springer-Verlag.

C.K. Riesbeck & R.C. Shank (1991), "From Training to Teaching: Techniques for Case-Based ITS".

E. Soloway (1992), "Interactive Learning Environments", ITS '92 Tutorial Notes.

Looi Chee Kit is a Senior Member of Technical Staff at Information Technology Institute, National Computer Board, 71 Science Park Drive, Singapore 0511.

BOOK REVIEW

by Justus Lewis

MANAGEMENT TOOLS FOR EDUCATIONAL MANAGERS

Allan Walker, Low Guat Tin, Chong Keng Choy and Kennett Stott. Published by Prentice Hall (1992)

Management Tools for Educational Managers, published in 1992 by Prentice Hall, is a collection of papers, contributed by Singaporean and Australian writers and edited jointly by Allan Walker, Low Guat Tin, Chong Keng Choy and Kennett Stott. Most of the Singaporean contributors are from Nanyang Technological University. The book has been written for "practicising educational managers who are committed to improving their knowledge and skills, ultimately for the betterment of their organisation. It is intended to be useful for educational leaders at all levels in the school, from heads of departments to principals."

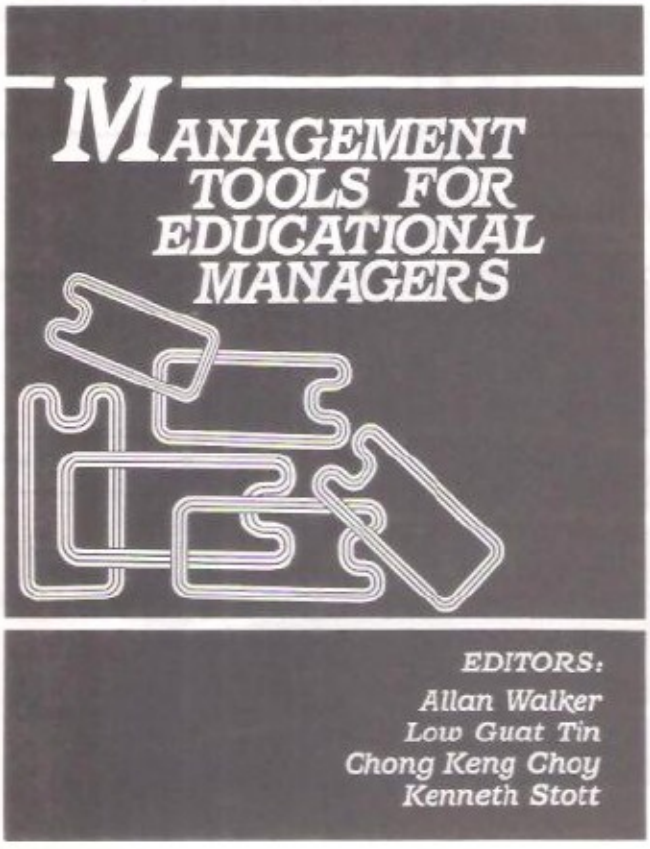
The papers cover a range of topics from individual self-improvement to strategic future planning for the organisation. They are: Self-Management, Using Management Time, Controlling Management Stress, Making Management Decisions, Management of Change, Marketing in Educational Institutions, Organising Group Tasks Effectively, Convenating for Excellent Executive Performance, Conducting a Needs Assessment and Futuring for Strategic Management of Educational Change.

Each chapter has a useful list of further references. The style of writing is straightforward and conversational. The content is copiously illustrated with case studies. The text is enlivened with cartoons, charts and tables. A number of activities and exercises are provided with each chapter, either for individual or group use.

This book would be useful to a number of groups, particularly those unfamiliar with recent thinking in management, leadership and personal growth and who want to obtain a quick grasp of something useful that could be immediately applied. Department heads and school principals would find it a helpful reference tool to have on the bookshelf. Consultants and trainers might find it a good source of suggestions for group activities in a variety of fields. Individuals looking for assistance in managing their time and reducing their stress levels would find a selection of productive techniques. In short, if you are looking for a quick overview of a variety of current management tools, and don't want to spend the time and money in extensive and systematic reading, this would be an excellent book to acquire. Its one obvious deficiency, which should be easily remedied, is the lack of an index.

Justus Lewis is Principal Education Development Officer in the Education Development Centre of Ngee Ann Polytechnic, Singapore.

Individuals looking for assistance in managing their time and reducing their stress levels would find a selection of productive techniques.



Inviting School Success for Everyone

Two lectures by

William Watson Purkey, Ed D

Lecture A: INVITING SCHOOL SUCCESS FOR EVERYONE

Lecture B: BECOMING AN INVITING EDUCATOR ... and living to tell about it

Date:4 July 1992
Time:9.00 am-

NOW AVAILABLE ON VIDEO

About the Speaker

Dr William Purkey is professor of Counselor Education at the University of North Carolina at Greensboro and Co-director, International Alliance for Invitational Education.

Dr Purkey has been awarded the *University of Florida Student Award for Instructor Excellence, the Good Teaching Award* by the Standard Oil Foundation, and the *Outstanding Teacher Award* by National Leadership Honour Society. He also received the *Distinguished Alumnus Award*, given by the School of Education, Lehigh University, and the 1989 *Professional Development Award* presented by the American Association of Counseling and Development.

Dr Purkey's interest is in inviting people to realise their full potential. He has written over 80 professional articles and six books, including *Self Concept and School Achievement*, now in its 18th printing, and *Inviting School Success*, now in its second edition. Two of his latest books are entitled *The Inviting Relationship* and *Invitational Teaching, Learning and Living*.

William Purkey on Video

Price: S$24.00 per copy (for members of ASCD Singapore)
S$30.00 per copy (for non-members)

Number of copies required: _____ @ S$_____ TOTAL S$_____

Name: _____

Address: _____

Tel: _____

Payment by crossed cheque only. Make cheque payable to "ASCD Singapore". Please mail your order and cheque to:

ASCD Secretariat
c/o Tele-temps Services
Blk 1002 Toa Payoh Industrial Park #06-1475
Singapore 1231

ASCD

ATTEND EDUCATION'S PREMIER PROFESSIONAL DEVELOPMENT EVENT!

48th Annual Conference and Exhibit Show

SHERATON WASHINGTON HOTEL, WASHINGTON, DC USA

March 27-30, 1993

INFORMATION FOR SINGAPORE PARTICIPANTS

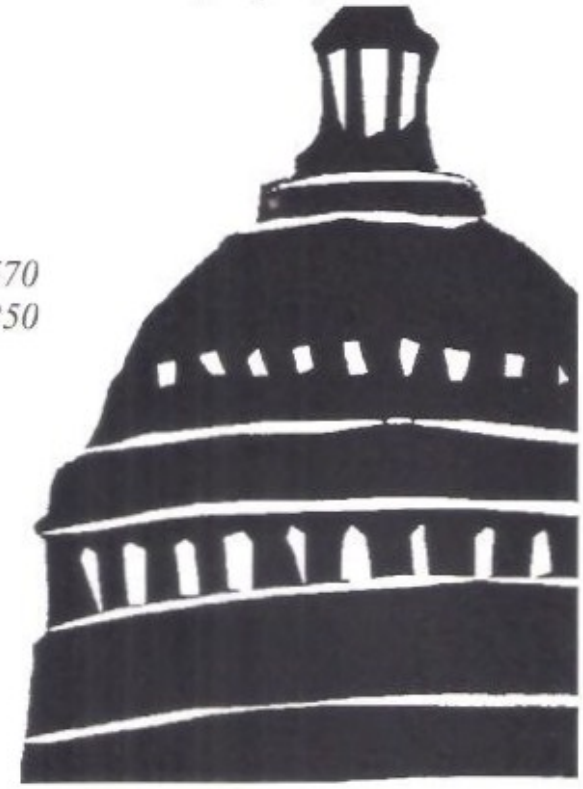
Estimated costs:

Airfare	*S$3,100*
Registration	*S$ 250*
Pre-Conference Institutes	*S$240-$570*
Hotel (per night)	*S$150-$250*

If you wish to attend the conference, contact:

Mrs Woo Yoke Yoong
Secretary
ASCD Singapore
Tel: 460 5478

by 31 December 1992



ISSN 0218-2491