issue **40** jan / feb 2013

All about Assessment

Big Idea

Assessment and Learning

Assessment is part of every teacher's job. But there's more to it than just finding out how much learning has taken place. Dr Jonathan Goh provides us with a broader perspective of assessment.

"Why is assessment needed?" asks Jonathan Goh, who is Associate Professor and Associate Dean (Research Support) in the Office of Education Research.

"Knowing how to do it is crucial to helping us know how well we have taught and how well the students have learned in school," he says. "It is paramount that teachers know how to assess students properly."

Proper assessment will let teachers have a clear picture of their students' true abilities and how much they have progressed over time. "There are, however, a lot of issues that teachers need to be aware of," adds Jonathan.

He guides us through the basics of assessment that every teacher must know.

The Role of Assessment in Learning

If we think of teaching and learning visually as a triangle, made up of curriculum, pedagogy and assessment, it often looks like a lop-sided one, notes Jonathan. This is because assessment is often given less attention than the other two components.

"Part of the reason, I suspect, is that people fear numbers," he says. "They think of it as statistics. As soon as you mention Means and Standard Deviations, they get really uncomfortable! Well, we need to understand that numbers are just numerical representations of qualitative meanings."

Assessment does indeed involve statistics. But it's not all about the numbers, as some teachers may fear.

More importantly, Jonathan asks us to consider these: What is assessment? Why do we need it? And how do we know how much our students have learned based on their test results?

> It is paramount that teachers know how to assess students properly.
>
> - Jonathan Goh
> *Office of Education Research*

NIE NATIONAL INSTITUTE OF EDUCATION SINGAPORE

An Institute of

NANYANG TECHNOLOGICAL UNIVERSITY

In this issue

Learning Beyond Assessments

Using Performance Tasks to Assess Learning

A Good Test

RESEARCH within REACH

Supported by **mc** Marshall Cavendish Education

> ## Formative assessment is useful for both teachers and students in modifying their practices to improve learning.
> – Jonathan *on the formative role of assessment*

Key Aspects of Assessment

Many people see assessment primarily as a way to *evaluate* skills or knowledge of our students. Jonathan believes assessment can do more than that.

"For a start, it's an opportunity for *enhancing* knowledge and skills through various modes of assessment, for instance, through projects, tests, assignments and presentations."

He lists three key issues of assessment that teachers need to think about:
1. How to plan for assessment
2. How to develop the test questions or instruments for testing
3. How to interpret the test results

Planning for assessment

To start things off, teachers should come together to draw up a curriculum plan. Almost like drawing up a map, they need to ask themselves what they want to achieve with assessment.

They then need to align the assessment with the curriculum. Teachers must decide on the spread of topics and difficulty levels that should be covered, and also what constructs they're really trying to test. All these would also help them in assessing the validity of their test.

"We can measure almost anything meaningfully," says Jonathan. "We just need to 'unpack' what (the variable) we are trying to measure, and ask questions on what we want to measure. In so doing, it will really help us better understand what we are trying to measure."

Is Your Test Valid?

Test validity is fundamental to the design of a good test. The three types of validity that teachers should look out for are:
1. *Face validity*: "If it's a Science test, we had better be sure that we have questions related to Science and nothing else!"
2. *Content validity*: "Test questions must be in accordance to what we have taught in the curriculum. It is also important to have a good spread of questions across the range of topics taught in class. Teachers need to also take into account of the difficulty levels of the questions asked in tests and exams.
3. *Construct validity*: "It is important that our test questions measure the psychological construct we intend to measure. I have come across many examples where the constructs shift. Here is a common one: Often, teachers use complicated or bombastic language in a Math problem-solving question. The child may get the answer wrong not because he or she didn't know the math but rather, the student was confused by the language used. Are we testing for English language or math ability?"

Read more about test validity in "A Good Test" in this issue.

Developing test questions

"Tests questions should not be convoluted," cautions Jonathan. "We should avoid words that are misleading or vocabulary that the students do not even understand."

But tests shouldn't be too easy either.

"You must have sufficient items and they should vary in terms of difficulty levels so that you have a good idea how the child is performing," he advises. Don't be afraid to "have questions that are difficult to really 'stretch' them". Only then can you know their true ability.

If we remember that the aim is to evaluate and enhance knowledge, we can develop test questions that enable further learning.

Interpreting the results

When it comes to interpreting results, invariably, the bell curve gets mentioned.

It does serve a purpose, says Jonathan. However, teachers need to understand three basic statistical concepts (Mean, Standard Deviation and Standard Error) to better understand how their students have performed.

Without a good grasp of these concepts, teachers may erroneously think that students are underperforming when their marks fall below the Mean, or celebrate when the class Mean for a particular subject is 80 out of 100.

Interpretation of test results gets more complicated when we make comparisons, either across time or levels, or between different students.

The scores alone do not tell us the full story. A student who scored 85 marks in Test 1 and then 75 marks in Test 2 may seem to be falling behind. But perhaps Test 2 was much harder than the first test.

The Formative Role of Assessment

Grades do matter, especially in high-stakes national exams, but educators also place much emphasis on formative assessment. "If you think about it, assessment is mainly formative because it's developmental," notes Jonathan. "Formative assessment is useful for both teachers and students in modifying their practices to improve learning, often through qualitative feedback."

Assessment reveal more than how much the students know. "It's also for us as teachers, how we're seeing growth and seeing how the students are learning the concepts we're teaching."

Assessment benefits not just students. It's also an opportunity for teachers to improve on their own pedagogies.

"We may look at the students' performances and reflect on how we teach. Then we adjust accordingly to make sure that these concepts are well understood by the students."

With the focus shifting towards 21st century competencies, the challenge for teachers today is not just to impart knowledge, but to create opportunities for students to show how well they have understood the concepts.

This may require teachers to create assessments that simulate real-life situations so that students can demonstrate learning through application.

Seen in this broader perspective, assessment can be formative for both students and teachers. Jonathan certainly thinks so.

"If teachers can see assessment as part of the whole teaching–learning process and continue to improve their assessment practices, it would certainly take our education system to new heights and the teachers to a new level of professionalism."

Jonathan Goh is the Guest Editor of this issue. He is an Associate Dean (Research Support) with the Office of Education Research and Associate Professor with the Policy and Leadership Studies Academic Group in NIE. His research interests include educational measurement, school leadership and management, student learning approaches and intercultural communication competence.

Research

Learning Beyond Assessments

As teachers, you put a lot of effort into preparing students for exams and in grading them. But in the process, learning may inadvertently take a back seat. Does assessment spell the end of learning?

Assessments are often equated with tests and exams. These are useful for ranking and placement purposes, when we need a practical way to make critical decisions.

However, a second—and perhaps more critical—purpose of assessment is to support and enhance learning. "In contrast to the regulatory approach, it's the educational purpose of assessment," says Assistant Professor Kelvin Tan.

The danger is when the regulatory functions dominate the educative benefits of assessment. "Once the assessment stops, the learning stops. Sometimes it signals the end of learning," he notes.

But this need not be the case. For Kelvin, it starts with a healthy view of assessment.

Assessment for Learning

"Mentally, everyone knows teachers cannot do the learning for the students. But in reality teachers act the opposite way," observes Kelvin. "They are teaching and assessing in a way as if it's all dependent on them."

So we labour over each test paper and exam script, making annotations and suggesting corrections. All that feedback was meant to help students improve on their learning. It was meant to feed back into their learning.

But more often than not, it doesn't. It doesn't get a chance to.

> Once the assessment stops, the learning stops. Sometimes it signals the end of learning.
>
> - Kelvin Tan
> *Curriculum, Teaching and Learning Academic Group*

Reference

Tan, K. H. K. (2012). Variation in teachers' conceptions of alternative assessment in Singapore primary schools. *Educational Research for Policy and Practice*. Advance online publication. doi: 10.1007/s10671-012-9130-4

Kelvin Tan is Assistant Professor with NIE's Curriculum, Teaching and Learning Academic Group. He lectures in the Management and Leadership in Schools programme in NIE for Heads of Department. His research interest is in assessment for learning.

In order not to let all that effort go to waste, Kelvin believes teachers should start passing the baton of learning back to the students.

We need to think more about what assessment should do for students' learning, and less about what students should be learning for assessment.

Assessment for learning prompts and enhances the quality of learning. "It happens when teachers introduce a notion of assessment that is integral to the students' learning, and not something that concludes learning."

Teachers' Conceptions of Alternative Assessment

This has prompted many teachers to think about an alternative to the current practices of assessment. But an alternative assessment practice may not always be a meaningful departure or an improvement from traditional assessment. Much depends on how teachers understand and use it in schools.

From his research, Kelvin found that most teachers are either conservative or pragmatic in their use of what they understand to be "alternative assessment".

Conservative teachers would rather not divert from the current practice. If there is some spare curriculum space, they may provide an additional worksheet or task. But this doesn't interfere with or distract from the existing traditional assessment.

Pragmatic teachers are also happy to abide within the existing curriculum boundaries and structures. If they find an alternative assessment practice to be effective, such as a learning journey project, they may give it 1 or 2 weeks within curriculum time.

Kelvin recommends taking a more progressive view of assessment, one that is integral to learning. It's about sustainable learning—beyond assessment, beyond their academic lives.

"Teachers who have this conception have a broad view of education. They are able to satisfy imminent pragmatic educational needs—scoring well in exams—and yet go beyond that."

A Progressive View of Assessment

Improving the quality of learning may require a paradigm shift, but it need not entail a drastic change in practices. The questions Kelvin ask is: What is alternative assessment an alternative to? What purpose do you want it to achieve?

The best assessments generate feedback that students can use. They are told how well they have done and what else they can do to improve their learning. This means giving students opportunities to act on that feedback.

For assessment to be effective, it must have an imminent impact on learning. This means students should act on the feedback sooner than later. This can be done by giving them a similar task soon after the previous task. Teachers can then track their progress.

"The underlying issue is how the curriculum is designed and understood. It's not so much the assessment," notes Kelvin.

Teachers can design the sequence of assessments to create a coherent series of tasks over a period of time, instead of a one-off task, such that each assessment task works in relation with the rest and they feed into each other.

"Design an assessment that generates feedback and provides a subsequent context for the feedback to be used," advises Kelvin. "This will create a dialogue between teachers and students on how and what to understand."

Alternative assessment need not mean additional work. "You decide what you have time for," says Kelvin. "How can students benefit from you? Once teachers can understand and appreciate that, they will rethink their teaching–assessment–feedback cycle."

Classroom

Using Performance Tasks to Assess Learning

Testing for knowledge recall is easy, but how can we tell if our students have really achieved deep understanding? This was what led the teachers at Raffles Girls' School (Secondary) to search for alternative assessment modes. They believe Performance Tasks provide this information.

With the introduction of the Raffles Programme, a 6-year Integrated Programme, at Raffles Girls' School (Secondary) (RGS) in 2006, the high-stakes O-level exams were done away with. With the traditional benchmark of academic performance gone, they looked for other measures to gauge learning.

"If you ask us what the goal of assessment is, we'll say it's mainly to gain evidence of learning," notes Mrs Mary George Cheriyan, Director of the RGS Pedagogical Research Lab (PeRL).

"It's not just the end product, or summative assessment. In order to use assessment to inform learning, formative assessment becomes important as well." Performance Tasks were one such assessment.

Evidence of Learning

RGS has been using Performance Tasks as an alternative form of assessment since 2006. As the name suggests, students are required to perform a task to demonstrate their understanding. All tasks have a real-world context.

As a form of alternative assessment, Performance Tasks provide evidence of learning in unconventional ways. Unlike traditional pen-and-paper tests, such tasks allow students to apply their learning in very practical ways.

For example, for an English Language task, students were asked to identify shop signs with linguistic errors. They then had to write a letter to the shop owner to suggest how to improve it.

Often, students are also given free rein as to how they can present their answers. One may choose to write a poem, another may produce a visual piece, while yet others may perform a skit. It's really up to them.

"When we give students the option to create the products in any form, they will go according to their preferred learning style," notes Mary. And this is exactly what they intended. "The reason why we went for alternative assessment is so that we can have multiple modes of learning."

Ms Lucille Yap, Mrs Mary George Cheriyan and Ms Choo Li Lin from Raffles Girls' School (Secondary).

If you ask us what the goal of assessment is, we'll say it's mainly to gain evidence of learning.

- Mrs Mary George Cheriyan
Raffles Girls' School (Secondary)

Using Performance Tasks

The RGS teachers discovered Performance Tasks when they went for a workshop on assessment principles conducted by the Alberta Assessment Consortium for teachers.

The appeal of Performance Tasks was enhanced by the fact that it is similarly emphasized in the Understanding by Design framework (UbD) framework, which RGS had already adapted for the school's curriculum design. Performance Tasks have been implemented school-wide.

"It is very important thing that you have custodians in the school. You must have that one person with oversight," says Mary.

At RGS, it is the Director of Academic Studies who vets all assessments and unit plans, a role Mary played until recently. All new teachers are also trained in how to design Performance Tasks.

Relying on a Rubric

Alternative modes of learning require alternative modes of assessment. Assessing the quality of learning is not an arbitrary process.

Performance Tasks rely on a carefully crafted rubric. For each task, the outcomes, criteria and the task itself are clearly detailed. Each task, regardless of subject, follows a fixed format.

This rubric guides the process of learning—it's not a one-off activity. Teachers spend time explaining the criteria, to ensure that students are clear about the expectations.

"It's transparent so the students are aware of how they are going to be assessed even before they actually attempt the task," explains Ms Choo Li Lin, who was herself a student at RGS and returned as an English Language teacher 5 years ago.

She adds, "With the traditional pen-and-paper assessment, what you see is a summative snapshot of the students' works. But with a Performance Task, we can structure it such that we see how the students progress over the whole learning process for that project."

In the English Language classroom, for example, this could mean giving them the opportunity to do multiple drafts. With each draft, the teachers can use the rubric to help them focus on particular areas that need improvement.

More online

See examples of Biology and History task rubrics online.

Designing the Task Rubric

Designing the rubric can be challenging, even for the more seasoned teachers, especially when the products of learning are so varied.

"Students are going to come up with all kinds of responses. Yet, the rubric needs to be able to capture these responses in a way that is quantifiable," says Mary. "How clear can you be? If you're so specific, then how can it be real-world?"

"The key challenge was the description," recalls Ms Lucille Yap, Head of Consultancy at PeRL. "How do we describe each standard with precision, so that we know what we are looking for and the students know what we are expecting from them?"

"We will always look through the perspective of the student: How would they understand what we've written?" explains Lucille, who teaches Geography. "We also check with one another frequently."

Li Lin believes the rubric adds to the learning process, and makes it meaningful for learners.

"Students are able to understand what they have done well, what they have not. It makes learning that much more quantifiable, and something they can also qualify; not just getting an A but why I got it."

The teachers at RGS have found that Performance Tasks help them to know if their students are really learning. It also helps students know how well they are learning.

"That comes back to the whole idea that assessment should also come with feedback," adds Mary. "Even in normal pen-and-paper testing, it's supposed to be used as a platform for feedback."

High Aspirations

If you're wondering where to begin, Mary recommends that you first ask yourself: What are your aspirations for your students?

For the teachers at RGS, they want their students to be competent and ready for real life. They believe that Performance Tasks help to achieve this learning outcome.

Looking at the evidence of learning in her classrooms, Lucille observes, "It is not the mere acquisition of knowledge and skills. They really can apply it to a higher level just through that performance task."

Mary also notes that the students are more engaged when they are able to learn in their preferred way. "I could see that it was meaningful to them because they are articulating their understanding in a way that they would want to."

"It was really very challenging when I first embarked on this," recalls Lucille. But a strong belief in the value of Performance Tasks keeps them going. "It's a journey," she says, "I'm still learning."

Mrs Mary George Cheriyan is Director of the RGS Pedagogical Research Lab (PeRL). **Ms Lucille Yap** is Head of Consultancy. **Ms Choo Li Lin**, who is a Teacher Specialist, is the newest addition to their research team.
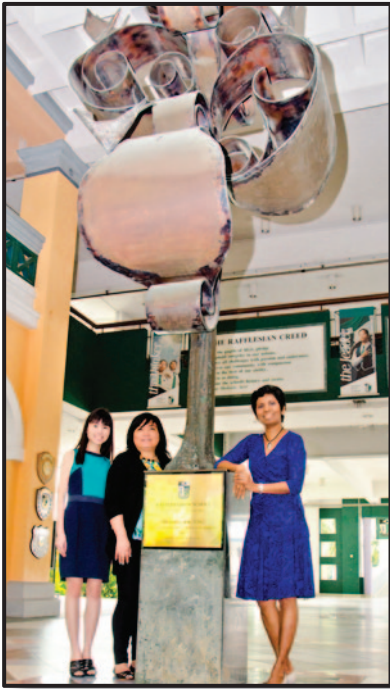
People

A Good Test

You've set the tests and graded the scripts. What next? Dr Lee Ong Kim explains the finer points of assessment and tells us what to look out for when testing our students.

Q: Why is it important to have a proper assessment system in schools?

We all know that the whole idea of teaching and learning in schools is for students' holistic development. This means that students should be assessed in all three domains—cognitive, affective and physical. Let's talk about cognitive assessment, since that is the main focus of schools.

Assessments should effectively inform teachers of students' progress, of their strengths, and of areas that are still not clearly understood. The assessment must therefore validly indicate to the teacher each student's problems in conceptual understanding of the material taught, so that a plan for remedial action may be made.

Also, students will have additional learning through the assessment itself, even if they do not perform well on it. Many talk about assessment of learning, for learning and as learning. But in schools, it is mainly for learning because it is invariably a formative assessment.

So a proper assessment system is needed to safeguard test integrity and validity in order to enhance the teaching–learning process.

Q: What is assessment validity?

There are several types of test validity—namely, face validity, content validity, construct validity, and criterion-related validity. *Face validity* is simple. As long as a test consists of questions on the subject it is claiming to be testing, then it has face validity. However, this is clearly not sufficient. There has to be content validity as well.

A test has *content validity* if the questions are on topics already taught to the students and on areas required by the curriculum. In addition, the levels of difficulty of the questions should be according to the test plan, that is, it must reflect the proportion of the number of questions planned for each level of the cognitive taxonomy.

Thirdly, the test must also have *construct validity*. This means that the questions must elicit the psychological constructs that they are purported to be testing. For example, if the test is to measure students' ability to use proper grammar in the English Language, then the questions must be on their grammatical skills. Test questions must not deviate from the intended purpose.

Sometimes we may find a question on a Math test that is based on a baseball game. The teacher must ensure that the question can be answered based on knowledge and ability on the mathematical construct intended and not based on knowledge of baseball games. Likewise, the language used on the Math test must not be at such a high level that the test construct shifts from math ability to language ability. This will make the test lose its construct validity.

Criterion-related validity is not so crucial for classrooms. In short, there are two types of criterion-related validity—concurrent validity and predictive validity. *Concurrent validity* means that if the test is said to be able to show how skilful the students are on one criterion, then performance on the test must also reflect the level of ability on the second, related criterion. *Predictive validity* would be the ability of the test to predict future success of the students, either on a job or at a higher level of study.

Q: How then should classroom tests be planned?

Classroom tests are usually planned through the drawing up of a "test blueprint", sometimes also referred to as the "table of specifications". The essential ingredients of the blueprint are the topics and sub-topics to be tested, the proportion of the number of questions on each sub-topic, the proportion of the number of questions at each level of the cognitive taxonomy for each topic and sub-topic, and the stated objective of each question.

A test blueprint is also very useful because if the teacher who is supposed to construct the test is unable to do so for some reason, another teacher will be able to take over the task without deviating much from what the first teacher would have set.

Q: What else is important in the assessment process for teaching and learning?

A test, no matter how well planned and prepared, will be rendered useless if the scores obtained by students cannot be properly interpreted. Testing is a measurement process where we measure the status of the students' learning at the time of testing. Measurement is always for the purpose of comparison.

Comparisons require an understanding of distributions with their Means and Standard Deviations, which can tell teachers the group status of their class compared to other classes, and how their students are spread out in their ability compared to other classes. Likewise, it is to compare how an individual student performs in comparison to others who took the test with him. Such interpretations of performance through comparisons are also termed *norm-referenced* interpretations.

Another form of performance interpretation is the *criterion-referenced* interpretation. This form of interpretation is like answering the question, "Is the student able to add two double-digit numbers that involves regrouping?" and other such questions. As long as the answer to such questions is "yes", the student has made the grade. Hence, a good teacher will be able

Assessments should effectively inform teachers of students' progress, of their strengths, and of areas that are still not clearly understood.

- Lee Ong Kim
Vice-President, World Educational Research Association

to make all (that is, 100%) of his or her students learn so well as to achieve a grade A. This is in contrast to using each student cohort as its own norm group, where even a group of A students will be spread out in order to get a certain percentage to be graded below A.

Q: What are some common errors teachers make when making comparisons?

Teachers are more likely to take test scores as "absolute", and conclude that a student who scored 86 marks is better than the one who scored 84 marks. This is a flawed interpretation if the Standard Error of the test is, say, 3 raw score points. Sometimes we even hear of students with an average of 92.5 marks being given a book prize and the student with an average of 92.3, who is not awarded anything, is interpreted as falling into "second" place. If we think about Standard Errors, it could well be that the so-called second place is in reality in first place and vice versa.

Another common wrong conclusion made by some teachers is to think that two classes with equal Mean scores are two equal groups. This is not necessarily the case because in one class the students may be more homogeneous while the other class may have a bigger spread of abilities about the same Mean. The class with the bigger spread has some weak students to be helped. The Mean of the class is the same simply because there are some better students who have "balanced off" the poor performances of the weaker students.

Comparisons also have to be made using proper scaling, which needs an understanding of scale linearization and measurement errors. But teachers need only know the basics of such issues—the basic statistics of Means, Standard Deviations and Standard Errors.

Q: How can a teacher tell if a student has improved over time?

This appears to be much harder for teachers to do in schools. This is because, to be able to compare students' performances across different tests at different levels or different time points, the tests will first have to be equated. Equating is required because the tests may not be of the same difficulty levels and hence equal scores on them do not reflect equal abilities.

Equating means putting the tests on a single common scale. We all know that to compare the lengths of two pieces of string, we should measure them on the same scale, the meter rule, for example. It would be really great if school teachers are taught how to equate tests so that their interpretations of students' growth may be made more accurately.

In fact, if teachers are to do research, such as to see which teaching method is better for a given subject or topic, the outcome variable being students' growth, it becomes all the more important that the tests be equated. Otherwise, the research conclusions will not be defensible.

Q: What assessment skills should teachers develop?

Teachers should develop skills in test planning, test-item construction, and interpretation of students' performance on the tests, in order to be able to take the next step in their lesson planning.

We should remember that the teacher's job consists of three main aspects—the curriculum, the pedagogy, and the assessment. It is not sufficient just to be able to interpret, plan and implement a curriculum, or to be excellent pedagogically.

Without excellent skills in assessment, the teacher will not know how well the students have learned, and they will also not know how well they have taught. Strength in all three aspects will increase the teacher's professionalism tremendously.

Lee Ong Kim is Vice-President of the World Educational Research Association (WERA). A former Associate Professor at NIE, his areas of expertise include measurement, evaluation, testing and quantitative research methods.

Online Exclusives

Diagnosing Misconceptions through Assessment

Where Anonymous Feedback is Welcome

Using Feedback to Enhance Learning

Learning in a Changing World