

# **Assessing skills and measuring performance of students on skills that are aligned with the needs of the 21st Century: A practical example**

**James Tognolini**

University of Sydney, Australia

**Abstract:** Malaysia is a multi-racial country comprises of diverse lifestyles, cultures as well as religions. This situation requires a different approach in developing the strong values of patriotism in order to preserve the harmony and strengthen the nation's identity of Malaysia. The construction of this patriotism framework relies on Malay culture and Islam as the official religion of the country as well as not disregards other ethnic cultures and religions. The goal of patriotism in the context of Malaysia leads to the enrichment of pride values as Malaysian, loyalty to the King and country, have a sense of belonging, has a level of discipline and obedience towards the laws of the country and becoming a competitive and productive society. In order to make this aspiration a reality, various attempts have been made in myriad number of schools, higher education institutes and education as well as community which is driven by the government, private organizations as well as NGOs. Apart from that, a detailed discussion on issues related to the challenges in establishing racial harmony which is a major factor in the formation of Malaysian patriotism framework as well as the suggestions to enforce patriotism among Malaysian nation are also incorporated.

## **1. Introduction**

The purpose of this paper is to outline some steps that could be taken to maximise the chances of producing comparable assessments across systems as they move towards developing an approach to assessing 21st century skills. The approach is a bottom-up approach where the intention is to empower teachers (and lecturers) in the process.

In introducing the reforms outlined above into education systems it is imperative that there be an overarching measurement theory that enables the various stakeholders (education systems and schools) to build their own assessments and measures while retaining the capacity to report against national standards.

Consequently, the first stage in the process involves articulating a theory of assessment predicated upon giving marks more common meaning by referencing them to standards and on a measurement model that will underpin the development of rubrics, the construction of “developmental continua”, etc.

### **The basic elements of standards-referenced systems**

A standards-referenced system is a model for giving meaning to achievement by referencing it to student learning or standards. This effectively shifts the focus in assessment from notions of

rank ordering students (comparing their performance to each other) to those of monitoring growth or progress and measurement along a developmental continuum (Bennett, et. al., (2012); Tognolini and Stanley, (2007)). It requires the articulation of what is meant by growth in a subject or construct. Rather than just a mark in an examination the system provides students with a description of the types of knowledge and skills that they have acquired in a subject at the end of a course.

When talking about assessment it is important to have a common understanding of some of the key terms: assessment, testing, evaluation and measurement. Inside and outside of education circles these terms are often used in overlapping and inconsistent ways.

Assessment involves professional judgment about student performance with respect to a continuum of development and is based upon the image formed of the student by the collection of evidence (Tognolini and Stanley, 2007).

Assessment is an inclusive term, which refers to all those processes used to collect information and make judgments about student achievement (Davidson and Tognolini, (2013)). Within each knowledge domain, teaching experience and subject expertise helps develop the image of achievement embodied in the standards. Testing is just one way of collecting information about students. As a formal process, it is a structured form of assessment collected according to specified procedures (question types, answer formats, etc.). Evaluation is when performance data is summarised by assigning a grade, comment or a mark and a judgement is made regarding the value of the image (it is good or bad; it is worth an A; it is a high distinction; etc.). Finally, measurement is the process of assigning a number to the performance to represent position with respect to the developmental continuum underlying the performance and indicates how much of the property (construct) being assessed is present (Davidson and Tognolini, (2013)).

Standards-referenced systems generally comprise a curriculum (syllabus or framework) that describes through its statement of aims, objectives, learning outcomes and content, what is developed and to be understood in an area of learning (Bennett, et. al., (2012)). Teaching and learning is based on the curriculum. The most important sources of information for the design of assessments and judging attainment of curriculum standards are the learning outcomes and content.

Performance (achievement) standards are explicit statements of student performance that describe the levels of achievement along the developmental path within the learning area.

The outcomes are developed to enable the students to achieve the performance standards and as such, show growth in relation to the construct being assessed. The performance of students as reflected through the assessment tasks (both formative and summative) is then referenced to these standards.

As stated previously, examples from the work being done at the University of Sydney in measuring graduate qualities will punctuate this paper. In the case of the University of Sydney, the graduate qualities (outcomes):

- describe what characterises learning within the University and the generic curriculum offerings and must be developed with these outcomes in mind; and,
- should be organised in a way that enables student achievement relative to the graduate qualities to be taught and measured.

When constructing assessment tasks, the marking rubrics (and options in the case of multiple-choice items) should reflect the theory. In a standards-referenced system, tasks (items or

---

questions) should be set in a way that provides evidence of where the students are located along the developmental continuum. If this is done then every response can be interpreted in terms of location (and hence ability, knowledge and skills) and should give an indication of what needs to be done to improve learning.

Some basic task development requirements would include making sure that

- the items and tasks (e.g. tests, assignments, practical work, and projects) are aligned to the content standards (outcomes) articulated in the syllabus;
- the items, and tasks that are developed enable students at different stages in their learning to demonstrate what they know and can do; and,
- a range of different tasks is used to generate a reliable and valid estimate of the student's location along the developmental continuum (Bennett, et. al., (2012)).

The contemporary interest in reporting against educational 'benchmarks' is based on standards referencing. Standards are defined in terms of more global descriptions of achievement and provide valuable information about the relative progress of student performance with respect to knowledge and skill development.

### **The developmental continuum**

Figure 1 gives a schematic representation of a developmental continuum. GQ1 to GQ 9 represent the graduate qualities (GQ1: Depth of disciplinary expertise, GQ2: Critical thinking and problem solving, GQ3: Communication (oral and written), GQ4: Information/ digital literacy, GQ5: Inventiveness, GQ6: Cultural competence, GQ7: Interdisciplinary effectiveness, GQ8: An integrated professional, ethical and personal identity, and GQ9: Influence).

The Levels 1, 2, 3, ... represent increasing quantities of the various graduate qualities; and, are descriptions of what it is students must know, be able to do and "value" to achieve the level. The descriptions are cumulative in the sense that to achieve Level 3, students have Level 2 PLUS some MORE of the property, construct or graduate quality, etc. These descriptions are referred to as the performance standards (Bennett, et. al., (2012)). It is a requirement of the measurement theory that underpins this approach that the Levels are cumulative.

The number of Levels can vary across graduate qualities and is determined by how many different categories (levels) can be explicitly described in a way that enables the assessors to be able to distinguish between levels of performance within the graduate quality. The aggregate of the performances across the graduate qualities comprises "university learning".

One of the key challenges is to be able to write the performance standards clearly and meaningfully for the students, lecturers and community; each of whom will use them in different ways ((Sadler, (2005)).

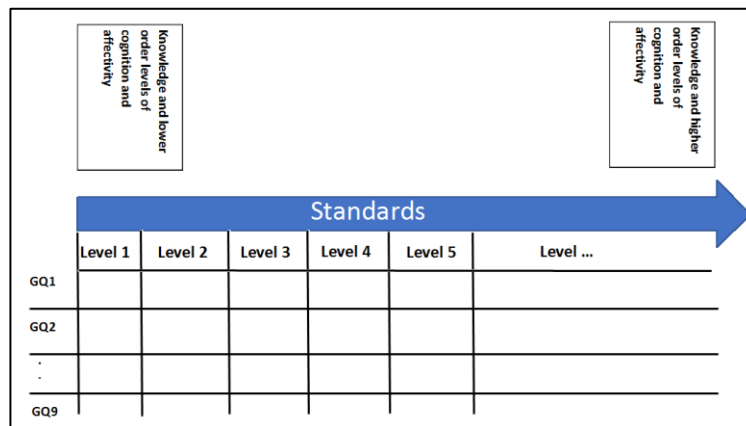


Figure 1: Schematic representation of growth of the Graduate Qualities (GQ1: Depth of disciplinary expertise, GQ2: Critical thinking and problem solving, GQ3: Communication (oral and written), GQ4: Information/ digital literacy, GQ5: Inventiveness, GQ6: Cultural competence, GQ7: Interdisciplinary effectiveness, GQ8: An integrated professional, ethical and personal identity, and GQ9: Influence) where increasing levels show increasing higher order of cognition and “affectivity”.

The requirements of the performance standards

The performance standards should

1. describe performance expectations and proficiency levels in the context of a clear conceptual framework and be built on sound models of student learning (developmental continuum) and affective domain development.
2. be clear, detailed, and complete; reasonable in scope; and both rigorous and well-grounded in the knowledge and affective domains.
3. be elaborated so that curriculum, teaching and assessment are all aligned.
4. facilitate the development of curriculum (and associated assessments) that will imbed the qualities across curricula.

Figure 1 is really an analytic marking rubric ((Sadler, (2005)) and the requirements of the performance standards are the same as those of marking rubrics used in assessment at the task level (Lasater, (2007)).

### Analytic marking rubrics

Analytic marking rubrics provide a guide to marking all types of performance based on how the students perform on the separate criteria (graduate qualities in this case) related to the task.

In the case of the graduate qualities, the qualities themselves are the criteria for the analytic rubric (See Appendix 1 as an example of a critical thinking rubric sourced from the University of Sydney).

The main advantage of analytic marking rubrics is that they convert performance into a score and in the case of a measurement model, a location on a scale. A second advantage is that they enable everyone (students, lecturers and community) to see what is required of them to achieve the various levels of performance on each of the criteria that comprise the task.

The main disadvantage is that they are more difficult to write because there are more criteria. A

second disadvantage is that consistency among the different markers is generally quite low.

It must be stressed that in most, if not all programs, student tasks are already assessed using analytic or holistic rubrics. There is however some variation in the extent to which these rubrics are articulated and the extent to which they meet the developmental requirements of a measurement model that enables them to be effectively evaluated.

One of the first challenges, is to develop in a consensus manner, rubrics (standards) for each of the graduate qualities that meet the requirements outlined above (Appendices 2 and 3 show the basic process to be used for building rubrics). These rubrics together will define the broader learning that characterises the curriculum.

### **Measurement of student performance against rubrics**

Building the rubrics with the intention of measuring student performance is the first stage of the measurement process. However, once the rubrics are built and validated they must be used to measure student performance ((Sadler, (2005); Webb, (2007)).

This process can be carried out in different ways at different levels within a system. At the University level, it would be possible to add some statements related to those qualities that address “values” to the Course Experience Questionnaire and ask graduates to indicate the extent to which they agree or disagree with the statement. This would give some baseline data and would enable cross-temporal comparisons at a macro level. The self-report measure used here has the same limitations as most self-report measures and as such it would be just one piece of information that would be used to generate some evidence that the University is having an impact on the graduate quality being assessed. Cultural competence is one graduate quality that lends itself to such an assessment.

Scenarios have been used to assess critical thinking. A similar assessment could be carried out at the university level on a sample of graduate students to provide some baseline data on performance on the graduate quality of critical thinking. Repetition of such an assessment with different samples across years would give some evidence as to the impact that the University is having on this graduate quality. The reliability and validity of such assessments would need to be determined. It is one approach in which the University can get an indication as to the impact of the efforts to incorporate the graduate qualities into the University programs.

In both these cases there would need to be standard setting exercises carried out to establish cut scores on the assessments that align to the performance standards of the University Graduate Quality rubrics and reporting would be done against the rubrics.

Perhaps the best way to ultimately measure performance of students against the University rubrics is to aggregate up the assessments against discipline specific rubrics from the unit level to the program level and finally to the University level. This is the most direct method of assessing performance against the Graduate Qualities and is based on the informed professional judgement of the lecturers themselves.

It may take some time as apart from building the rubrics at the different levels, there is a need to provide professional development for the teachers who will have to assess against the rubrics in a comparable and fair way. This will also involve supporting the lecturers in developing assessment tasks that enable the students to demonstrate performance on the rubrics across the

length of the program.

## Conclusion

Many systems have identified graduate qualities (generic skills or 21st century skills) in their strategic plans. Few have committed to measuring performance against these qualities. The reality is that if there is no attempt to measure impact, then the chances of these qualities being implanted into the learning environment of the students is doubtful.

In undertaking to measure student performance on these graduate qualities the University of Sydney has committed itself to the bold task of not only making sure that the measures that are used to report performance against the graduate qualities are as valid and reliable as possible, but also that the graduate qualities themselves are incorporated into the curriculum and assessments of the units that are used to transition knowledge skills and behaviours to its students.

To be successful it requires the support of the teaching staff. In addition, the process needs to be founded on a sound measurement model that will maximise the chances of students being assessed on qualities that are critical to success in the 21st century in a fair and consistent way. The same measurement principles can also be used to provide optimal processes for effective assessment practice in collaborative and project learning settings; and, streamline and reduce summative assessment at the unit of study level.

## References

- Bennett, J., Tognolini, J., Pickering, S. (2012). Establishing and applying performance standards for curriculum-based examinations. *Assessment in Education: Principles, Policy and Practice*, 19(3), 321-339.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, NY: Longmans, Green.
- Davidson, M. and Tognolini, J. (2013). *Assessment, standards-referencing and standard setting*. In Magdalena Mo Ching Mok (Ed.). *Self-directed Learning Oriented Assessments in the Asia-Pacific*. Springer Science+Business Media, Dordrecht.
- Lasater, K. (2007). Clinical judgment development: Using simulation to create an assessment rubric. *Journal of Nursing Education*, 46, 496-503.
- Sadler, D.R. (2005). *Interpretations of criteria-based assessment and grading in higher education*. *Assessment & Evaluation in Higher Education*, 30, 175-194.
- Tognolini, J. and Stanley, G. (2007). Standards-based assessment: A tool and means to the development of human capital and capacity building in education. *Australian Journal of Education*, 51 (3), 129-145.
- Webb, N. L. (2007). Issues related to judging the alignment of curriculum standards and assessments. *Applied Measurement in Education*, 20, 7-25.

## Appendix 1: Definition, components and rubric for critical thinking and problem solving

### Critical Thinking and Problem Solving

#### Definition

Critical thinking and problem solving are the questioning of ideas, evidence and assumptions in order to propose and evaluate hypotheses or alternative arguments before formulating a conclusion or a solution to an identified problem.

#### Components

- Definition of problem or issue in context
- Critical questioning of ideas, evidence and assumptions
- Creation and evaluation of hypotheses or alternative arguments
- Formulation of defensible conclusions and best possible solutions.

**Appendix 1:** Definition, components and rubric for critical thinking and problem solving (continued)

<b>Graduate Quality: Critical Thinking and Problem Solving</b>				
<b>Components/Indicators</b>	<b>0 = Bottom level performance indicator</b>	<b>1</b>	<b>2</b>	<b>3 = Top level performance indicator</b>
Definition of problem or issue in context	Does not offer a definition of the problem or issue, makes no attempt to situate problem or issue in context or to explain why it matters, makes no attempt to identify criteria against which to evaluate various solutions.	Provides a superficial definition of the problem or issue, shows that the problem or issue is situated in a context, but is not yet able to show why the context matters, or to identify criteria against which to evaluate various solutions.	Provides a useful definition of the problem or issue, shows that the problem or issue is situated in a context, shows understanding of some details of that context and can explain why these matter, provides definitions of some of the key terms, can identify some desirable features of various solutions.	Insightful and articulate. Analyses and understands a context by consulting a suitably broad range of informational sources, identifies and appropriately frames a problem or issue within that context, phrases the problem or issue clearly in their own words, defines key terms, explains why this problem matters, sets out criteria against which to measure various solutions.



**Appendix 1:** Definition, components and rubric for critical thinking and problem solving (continued)

<b>Graduate Quality: Critical Thinking and Problem Solving</b>				
<b>Components /Indicators</b>	<b>0 = Bottom level performance indicator</b>	<b>1</b>	<b>2</b>	<b>3 = Top level performance indicator</b>
Critical questioning of ideas, evidence and assumptions	Defers to received opinion without evaluating sources of information and without considering possible bias and error, does not recognise genuine expertise, does not identify or question methodologies used, is swayed by mere rhetoric, does not engage with evidence, does not consider other historical, intercultural or interdisciplinary perspectives.	Recognises that ideas, evidence and assumptions need to be examined, makes an attempt to identify possible bias and error, recognises that not every self-described expert has genuine expertise, identifies and avoids at least some fallacious rhetoric, provides evidence to justify conclusions.	Questions received ideas, evidence and assumptions, engages with the work of genuine experts, critiques fallacious rhetoric, engages in rational argument, assesses currently available evidence, shows awareness that new evidence may be discovered, shows awareness of differences in perspective.	Open-minded and intellectually rigorous. Critically examines received ideas, evaluates the credibility and methodology of authorities and experts, distinguishes sound reasoning from mere rhetoric, assesses currently available evidence, engages with competing views from various historical, intercultural and interdisciplinary perspectives, locates new evidence.

**Appendix 1:** Definition, components and rubric for critical thinking and problem solving (continued)

<b>Graduate Quality: Critical Thinking and Problem Solving</b>				
<b>Components/Indicators</b>	<b>0 = Bottom level performance indicator</b>	<b>1</b>	<b>2</b>	<b>3 = Top level performance indicator</b>
Creation and evaluation of hypotheses or alternative arguments	Does not develop original hypotheses or arguments. Does not see how various hypotheses could be tested.	Recognises that current hypotheses and arguments may be suboptimal and ought to be evaluated, assesses at least some of the existing hypotheses and arguments.	Generates some new hypotheses and arguments, shows awareness of how they could be compared and tested, carries out at least some of these tests.	Creative and judicious. Generates original hypotheses and arguments. Figures out how to test relevant hypotheses and arguments via reasoning, observation, or experiment, carries out these tests, evaluates the results.
Formulation of defensible conclusions and best possible solutions	Does not offer any solution or conclusion, or dogmatically asserts a solution without being able to defend it. Does not identify and apply the relevant evaluative criteria.	Formulates an incomplete or unconvincing solution or conclusion, and is not yet able to offer a convincing defence with reference to the relevant evaluative criteria.	Offers a solution or conclusion based on engagement with some of the relevant evidence, defends this solution or conclusion in light of relevant evaluative criteria.	Wise and decisive. Decides on the balance of the evidence, formulates conclusion or solution clearly in their own words, identifies the proper scope and significance of the conclusion commensurate with methods used, explains why this conclusion or solution is best when measured against relevant evaluative criteria.