

13th Annual

Thinking & Learning Conference

DR JANELLE WILLS

Friday 20 May

**Collaborative Teams that Transform
Assessment: The next steps in PLCs**

Session 1

MELBOURNE

DR JANELLE WILLS

Dr Janelle Wills, PhD, is the director of Marzano Institute Australia. She is the lead training associate for High Reliability Schools, The Art and Science of Teaching and other Marzano Research topics. She works extensively with schools, regions and systems throughout Australia.



With over 30 years of teaching and leadership experience, Dr Wills maintains a strong commitment to continued learning that enables her to remain both informed and innovative in her approach. Throughout her career, she has been adept at linking theory with practice, resulting in the development of significant initiatives both within schools and at a sector level. Dr Wills firmly believes in the importance of teaching as a profession and fervently promotes the need for teachers to actively engage with research through action research and reflective practice.

Dr Wills's PhD thesis focused on self-efficacy and contributed to multiple fields of knowledge, including special education, gifted education, assessment and feedback.

A message from Hawker Brownlow Education

We hope that you have found these conference papers and the accompanying sessions useful. Please be aware that the contents of these papers are the intellectual property of the speaker and no reproduction for any purpose is authorised. We urge you to take care of this booklet. Replacement copies will not be made available either during or after this conference.

Published in Australia by



This handout was created by Hawker Brownlow Education for the proceedings of the Hawker Brownlow 13th Annual Thinking & Learning Conference. All rights are reserved by Hawker Brownlow Education. It is a violation of copyright law to duplicate or distribute copies of this handout by any means for any purposes without prior permission in writing from Hawker Brownlow Education. Professors and workshop presenters must first secure written permission for any duplication rights. For copyright questions, permission requests, or information regarding professional development contact:

Hawker Brownlow Education
P.O. Box 580, Moorabbin, Victoria 3189, Australia
Phone: (03) 8558 2444 Fax: (03) 8558 2400
Toll Free Ph: 1800 334 603 Fax: 1800 150 445
Website: www.hbe.com.au
Email: orders@hbe.com.au

© 2016 Hawker Brownlow Education
Printed in Australia

CODE: MELJW0101
0516

CHAPTER 4

Transforming Assessment

Once proficiency scales are in place, a powerful answer becomes possible to the question, How do we know if our students are learning? Indeed, proficiency scales can transform classroom assessments into tools to determine how much students have learnt as well as their current status at a particular point in time. There are four concrete steps collaborative teams can take when designing and using assessments based on proficiency scales.

1. Use proficiency scales as the basis for all assessments.
2. Design an assessment blueprint.
3. Write the assessment items.
4. After administering the assessment, score it and discuss the results.

Here we discuss each step in more detail.

Using Proficiency Scales as the Basis for All Assessments

As described in chapter 3 (page 31), a proficiency scale is basically a set of objectives on a specific topic organised into a progression representing their level of difficulty. This makes them the perfect framework from which to design assessments. To illustrate, consider the proficiency scale in figure 4.1 (page 50). In this scale, there are five objectives at score 2.0, one objective at score 3.0 and one objective at score 4.0. Recall from the discussion in chapter 3 (page 31) that some educators prefer to avoid the use of the verb *understand*. We, however, believe it is quite appropriate as long as teams provide more specificity regarding expectations of students elsewhere – for example, in student-friendly scales or instructional activities.

It is not uncommon to design an assessment that addresses all three levels of content in a scale. In such cases, individual items are designed for score 2.0, 3.0 and 4.0 content. This is frequently done with pre-assessments – those administered before instruction has begun. However, throughout a unit, teams might design one or more assessments that are focused on one level of content only. For example, at the beginning of a unit of instruction, a few assessments might focus only on the score 2.0 content. Later on in the unit, assessments might focus only on score 3.0 content. Here, the teacher assesses students to the level of content

COLLABORATIVE TEAMS THAT TRANSFORM SCHOOLS

that has been explicitly taught in class. If a teacher has only taught the content at the 2.0 level, it might be appropriate to assess only to the 2.0 level. However, some students may be beyond the level at which the teacher has instructed, so providing opportunities for students to demonstrate that understanding may also make sense. Score 4.0 is an example of this. By definition, score 4.0 goes beyond what was explicitly taught, so teachers will commonly include score 4.0 assessment items even though very little instruction has taken place for that level. Score 4.0 assessment items do not count against a student; they provide an opportunity for students to demonstrate a higher level of understanding.

Score 4.0	Students will identify two competing claims about a text, support each with textual evidence and decide which of the claims is better supported.	
	Score 3.5	In addition to score 3.0 performance, partial success at score 4.0 content
Score 3.0	Students will make claims about what a specific text says explicitly and use relevant textual evidence to support those claims.	
	Score 2.5	No major errors or omissions regarding score 2.0 content and partial success at score 3.0 content
Score 2.0	Students will understand the concept of an inference. Students will understand the concept of a claim that is supported by evidence. Students will understand the concept of evidence explicit in a text. Students will find or recognise claims that are supported by textual evidence provided by the teacher. Students will find or recognise textual evidence to support claims provided by the teacher.	
	Score 1.5	Partial success at score 2.0 content but major errors or omissions regarding score 3.0 content
Score 1.0	With help, partial success at score 2.0 content and score 3.0 content	
	Score 0.5	With help, partial success at score 2.0 content but not at score 3.0 content
Score 0.0	Even with help, no success	

Figure 4.1: Sample proficiency scale.

Designing an Assessment Blueprint

For each common assessment, the collaborative team creates an *assessment blueprint* (Marzano & Yanoski, 2016), outlining which types of items and how many items there will be on the assessment. To do so effectively, members of the collaborative team must first discuss the content at each level of the scale in order to ensure a strong match between the assessment items and the content. For example, if the content at score 3.0 of a scale addresses students' ability to write an informative essay with special attention to the transitions between paragraphs, assessment items consisting of short-answer and essay questions are more appropriate than a multiple-choice test. Table 4.1 compares four types of assessment items to the proficiency scale levels they are most appropriate for assessing.

Transforming Assessment

Table 4.1: Assessment Item Types and Levels of Knowledge

Types of Assessment Items	Levels of Knowledge		
	Score 2.0	Score 3.0	Score 4.0
	Basic knowledge and skills that students have learnt during the instructional unit – fairly easy	More complex knowledge and skills that students have learnt during the instructional unit – doable if students were paying attention	Inferences or applications that go beyond what students were explicitly taught – challenging
Forced-Choice Items	Short items with a small number of correct responses; options are often included (for example, multiple choice, matching, alternative choice, true/false, fill-in-the-blank, multiple response [asks for two or more correct answers])		
Short Written Response	Items that require the construction of one to a few sentences		
Essay		Longer written response of several paragraphs; covers more information and often requires students to connect, analyse or apply information; usually requires students to use multiple levels of knowledge	
Oral Response	Spoken version of forced-choice items or short written responses; longer spoken items such as question-and-answer sessions or structured discussion		

Source: Adapted from Marzano, 2006.

Once the item types are determined, teachers can then consider how many items are needed. It is important that they develop an adequate number of assessment items for each objective in a scale. Each level of the scale will likely have a different number of assessment items associated with it, due to the differing types of content. Score 2.0 assessment items are usually shorter and each cover one or two specific facts or details, so there may be more items at this level than at score 3.0 or 4.0, where each item may cover several aspects of the content and require a longer answer. In general, in a comprehensive assessment, score 2.0 content has five or more items, score 3.0 content has two or more items, and score 4.0 content might only have one or two complex items (Marzano & Yanoski, 2016; Marzano et al., 2013). For example, an assessment might consist of 10 multiple-choice questions (score 2.0), three short constructed-response items (score 3.0) and one essay question (score 4.0).

Writing the Assessment Items

Given the preparation described in the preceding sections, the actual writing of an assessment should be relatively straightforward. To illustrate, consider the assessment in figure 4.2 (pages 52–53), which is based on the proficiency scale about making claims in figure 4.1.

COLLABORATIVE TEAMS THAT TRANSFORM SCHOOLS

SECTION 1

1. When we say that we have made an inference about a text, we mean that we have noticed something that is:
 - A. Directly stated
 - B. Indirectly hinted at
 - C. Explained in a footnote
 - D. Not present in the text at all
2. If you are writing an essay about a book and want to support your claim with textual evidence, the best thing to do would be:
 - A. Cite a direct quote
 - B. Paraphrase the text
 - C. Refer to a quote from an expert
 - D. Either A or B
3. Consider the following quote from *To Kill a Mockingbird* and then select the claim that it best supports: "Mockingbirds don't do one thing but make music for us to enjoy. They don't eat up people's gardens, don't nest in corncribs, they don't do one thing but sing their hearts out for us. That's why it's a sin to kill a mockingbird" (Lee, 1960, p. 119).
 - A. Children are often smarter than adults expect.
 - B. Punishing innocent people is wrong.
 - C. Some animals are pests.
 - D. Always stand up for what you believe in.
4. Consider the following claim about *To Kill a Mockingbird* and select the quote that best supports it: In Maycomb, being masculine or manly means being physically able.
 - A. "For some reason Dill had started crying and couldn't stop; quietly at first, then his sobs were heard by several people in the balcony" (Lee, 1960, p. 265).
 - B. "Jem grabbed his left wrist and my right wrist, I grabbed my left wrist and Jem's right wrist, we crouched, and Dill sat on our saddle. We raised him and he caught the window sill" (Lee, 1960, p. 70).
 - C. "Jem was scarlet. I pulled at his sleeve, and we were followed up the sidewalk by a philippic on our family's moral degeneration, the major premise of which was that half the Finches were in the asylum anyway, but if our mother were living we would not have come to such a state" (Lee, 1960, p. 136).
 - D. "Our father didn't do anything ... Atticus did not drive a dump-truck for the county, he was not the sheriff, he did not farm, work in a garage, or do anything that could possibly arouse the admiration of anyone" (Lee, 1960, p. 118).
5. Consider the following claim about *To Kill a Mockingbird* and select the quote that best supports it: Women in the story are typically polite on the outside but cruel underneath.
 - A. "I wondered at the world of women ... I must soon enter this world, where on its surface fragrant ladies rocked slowly, fanned gently, and drank cool water. But I was more at home in my father's world. People like Mr. Heck Tate did not trap you with innocent questions to make fun of you; even Jem was not highly critical unless you said something stupid" (Lee, 1960, pp. 312–313).
 - B. "I felt the starched walls of a pink cotton penitentiary closing in on me, and for the second time in my life I thought of running away" (Lee, 1960, p. 182).

Transforming Assessment

C. "When we arrived at the Landing, Aunt Alexandra kissed Uncle Jack, Francis kissed Uncle Jack, Uncle Jimmy shook hands silently with Uncle Jack" (Lee, 1960, p. 107).

D. "Miss Caroline was no more than twenty-one. She had bright auburn hair, pink cheeks, and wore crimson fingernail polish. She also wore high-heeled pumps and a red-and-white-striped dress. She looked and smelled like a peppermint drop" (Lee, 1960, p. 21).

SECTION 2

1. Examine the following three quotes from *To Kill a Mockingbird* and make a claim that is supported by all three. Then, explain how each quote supports your claim.

"Miss Caroline seemed unaware that the ragged, denim-shirted and floursack-skirted first grade, most of whom had chopped cotton and fed hogs from the time they were able to walk, were immune to imaginative literature" (Lee, 1960, p. 22).

"In Maycomb, if one went for a walk with no definite purpose in mind, it was correct to believe one's mind incapable of definite purpose" (Lee, 1960, p. 199).

"[Atticus] did not do the things our schoolmates' fathers did: he never went hunting, he did not play poker or fish or drink or smoke. He sat in the livingroom and read" (Lee, 1960, p. 118).

2. Make a claim about the way most citizens of Maycomb treat children and the way Atticus treats children. Find at least two pieces of textual evidence to support your claim.

3. Make a claim about a theme or point that Harper Lee conveys through the story of Tom Robinson's arrest and trial. Use at least three pieces of textual evidence to support your claim.

SECTION 3

1. Make two opposing claims about a theme, character, relationship or other situation in *To Kill a Mockingbird*. Support each claim with at least two pieces of textual evidence and then explain which claim is better supported.

Figure 4.2: Sample assessment at three levels.

A key criterion when writing assessment items is *validity*, which means that items measure what they are intended to measure. Using proficiency scales as the basis for writing assessment items helps ensure validity because the scale delineates both the content and its level of difficulty. If a collaborative team uses a proficiency scale to create items based on the content articulated at each level of the scale, each item should address one level of the scale only.

When formatting the assessment as it will be presented to students, we recommend separating 2.0 items, 3.0 items and 4.0 items into distinct sections of the assessment. This is depicted in figure 4.2 – section 1 contains the items for level 2.0 content, section 2 contains the items for level 3.0 content, and section 3 contains the item for level 4.0 content. This will simplify the scoring process. If items at different levels are mixed, then teachers must keep track of the difficulty level of each item.

Finally, we should note that designing common assessments based on proficiency scales does not mean that previously constructed assessments need to be discarded. While it is certainly possible that a collaborative team would write entirely original items for an assessment, it is also plausible that they would take items from existing assessments, modifying them as necessary. Teachers can also *backmap* existing assessments by identifying the proficiency scale and level to which each item of an existing assessment relates (for more on backmapping, see Heflebower et al., 2014).

COLLABORATIVE TEAMS THAT TRANSFORM SCHOOLS

Scoring the Assessment and Discussing the Results

Once students have completed an assessment, the teacher should score it as soon as possible. Typically, scoring is done by individual teachers but it could also be done collaboratively if a team so chooses. It is extremely helpful if, when creating the assessment, a collaborative team also creates a scoring guide for that assessment. Scoring guides are particularly useful for items that are scored using multiple points. For example, items that are assigned five or more points should have a brief description of what type of answer receives a score of 5, what type of answer receives a score of 4, and so on. Tammy Heflebower, Jan K. Hoegh and Phil Warrick (2014) stated that scoring guides are advantageous because they

ensure fairness in assessment practices, provide more reliable interpretations of assessment information, and allow for more consistency in . . . scoring. These guides also help each teacher understand which items assess which proficiency level and how to identify correct, partially correct, and incorrect responses. (p. 50)

Teachers can also show scoring guides to students prior to the assessment to clarify the expectations for each level of knowledge.

There are two basic approaches to scoring an assessment that contains items at score 2.0, 3.0 and 4.0 levels of a proficiency scale: (1) using percentage scores at each level, and (2) using response codes at each level. After scoring a common assessment, collaborative teams should convene to analyse the results.

Using Percentage Scores

To illustrate the percentage approach, consider the example in table 4.2.

Table 4.2: The Percentage Approach to Scoring Assessments

Section	Item Number	Possible Points per Item	Obtained Points per Item	Section Percentage
Score 2.0	1	5	5	22/25 = 88%
	2	5	4	
	3	5	3	
	4	5	5	
	5	5	5	
	Total		25	
Score 3.0	6	10	7	15/30 = 50%
	7	10	4	
	8	10	4	
	Total		30	
Score 4.0	9	10	1	3/20 = 15%
	10	10	2	
	Total		20	

Transforming Assessment

The assessment scored in table 4.2 includes 10 items – five at the score 2.0 level, three at the score 3.0 level and two at the score 4.0 level. Each item has a specific number of points that students can possibly earn (third column). The fourth column reports the number of points a specific student earned on each item. The fifth column displays the section percentage, computed by dividing the obtained points by the possible points.

In table 4.2, the student acquired 88 per cent of the possible points for the score 2.0 level, 50 per cent of the points for the score 3.0 level and 15 per cent of the points for the score 4.0 level. Examining the overall pattern, the teacher then determines how well the student performed overall in reference to the scale. This is done by making decisions about the student's proficiency moving from score 2.0 through score 4.0. The score 2.0 percentage is 88 per cent, so the teacher concludes that the student obtained at least a score of 2.0 on the assessment. Next, the student's percentage score for the 3.0 content was 50 per cent. The teacher concludes that this is not enough to warrant an overall score of 3.0, but it is enough to warrant a score of 2.5. The teacher stops at this point. If a student has not provided enough evidence to warrant a score at one level, then they are not scored at the next level up.

Using Response Codes

With this approach, each student's response on each item is coded as *correct*, *partially correct* or *incorrect*, as opposed to assigning points to each item. For more specificity, teachers can use *high partial* and *low partial* in place of *partially correct*. After scoring individual items, the teacher determines the pattern of responses and assigns a score accordingly. For example, if a student's answers are correct on all items of the score 2.0 section of the test, partially correct on two items of the score 3.0 section of the test and correct on the third item of the score 3.0 section, and incorrect on the two items of the score 4.0 section of the test, that student would receive a score 2.5 (DuFour & Marzano, 2011). Table 4.3 displays this pattern of responses.

Table 4.3: The Response Codes Approach to Scoring Assessments

Section	Item Number	Correct, Partially Correct or Incorrect?	Section Pattern
Score 2.0	1	C	Correct
	2	C	
	3	C	
	4	C	
	5	C	
Score 3.0	6	PC	Partially correct
	7	C	
	8	PC	
Score 4.0	9	I	Incorrect
	10	I	
Overall Score			2.5

Source: Adapted from Marzano, 2010.

COLLABORATIVE TEAMS THAT TRANSFORM SCHOOLS

It is important to note that if the assessment addresses more than one proficiency scale, students will receive one score per scale, rather than one overall score (DuFour & Marzano, 2011). That is, if an assessment includes items that cover two different topics and was designed using two proficiency scales, a student might receive an overall score of 2.5 for one topic and an overall score of 3.0 for the second topic. The teacher does not assign an overall score for the entire test.

Analysing Results

After giving and scoring an assessment, teachers should discuss the results in a collaborative team meeting. The team might discuss questions such as the following.

- On which parts of the assessment did students perform well?
- On which parts of the assessment did students struggle?
- Were there any patterns evident in the student responses that we should discuss as a team?
- Which students are in need of special attention?
- Does the assessment need revision? Which items? Why?

In the next chapter, we address how answers to questions such as these translate into decisions regarding instructional planning. As indicated by the last question in the list, it may be necessary to revise an assessment for future use. When examining assessment results, it may become clear that a particular item is problematic; for example, if numerous students respond incorrectly on the same score 2.0 item but answer most or all of the 3.0 items correctly, the score 2.0 item should be re-examined. When this is the case, a collaborative team should remove the item or revise it to make it more consistent with the other score 2.0 items.

Using Multiple Types of Assessments

Proficiency scales also provide an opportunity to transform our perceptions of what constitutes an assessment. Technically, *assessments* are “planned or serendipitous activities that provide information about students’ understanding and skill in a specific measurement topic” (Marzano, 2006, p. 35). *Measurement*, on the other hand, is the act of translating the information about students gleaned from assessments onto some scale (Marzano, 2006). The use of a proficiency scale as a consistent basis for measurement allows teachers to be flexible in the types of assessments they use. Specifically, there are three broad categories of assessments, all of which are effective means of gathering information about student learning and translating the information to a scale. (For a comprehensive discussion of the three categories of assessments, see Marzano, 2010.) First, there are *obtrusive assessments*, in which instruction stops and assessment occurs. These typically take the form of traditional pencil-and-paper tests like figure 4.2 (pages 52–53), but could also involve oral exams or demonstrations. Second, there are *unobtrusive assessments*, in which instruction or normal classroom activity does not stop, but the teacher observes a student or students and records an assessment score. Often, the students are not aware that they are being observed. The third type of assessment is *student-generated assessments*. With these types of assessments, individual students decide how they will demonstrate a level of proficiency and take responsibility for doing so.

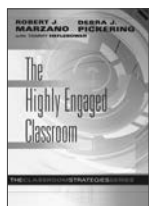
Using all three types of assessments provides teachers with great flexibility. Teachers do not have to administer the same number or the same type of assessments to all students. If a teacher is fairly confident that

RELATED RESOURCES

Available from Hawker Brownlow Education

Qty	Code	Title	Price
	MRL2779	A Handbook for High Reliability Schools	\$29.95
	MRL2861	A School Leader's Guide to Standards-Based Grading	\$27.95
	SF114041	ASCD Arias Publication: Self-Regulated Learning for Academic Success	\$15.00
	104017	Building Background Knowledge for Academic Achievement	\$29.95
	MRL6741	Coaching Classroom Instruction: The Classroom Strategies Series	\$39.95
	MRL7484	Collaborative Teams That Transform Schools: The Next Step in PLCs	\$35.95
	LSM5626	Essentials for Achieving Rigour Series: Creating & Using Learning Goals & Proficiency Scales: How Teachers Make Better Instructional Decisions	\$25.95

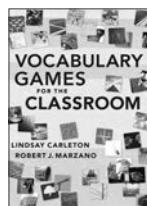
	MRL2878	Proficiency Scales for English and Mathematics Standards	\$38.95
	MRL2496	Questioning Sequences in the Classroom: The Classroom Strategies Series	\$29.95
	107001	The Art & Science of Teaching: A Comprehensive Framework for Effective Instruction	\$23.95
	MRL7634	The Highly Engaged Classroom: The Classroom Strategies Series	\$35.95
	HB7477	Transformative Collaboration: Five Commitments for Leading a Professional Learning Community	\$39.95
	MRL1222	Vocabulary For The Common Core	\$39.95
	MRL6217	Vocabulary Games for the Classroom	\$40.00
Total (plus freight) \$			



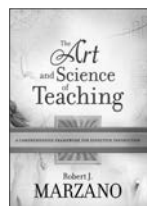
MRL7634



SF114041



MRL6217



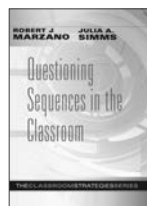
107001



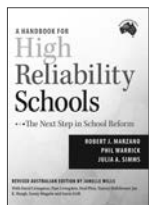
104017



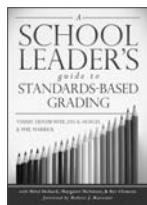
MRL6741



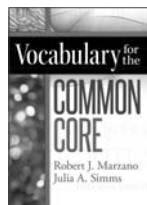
MRL2496



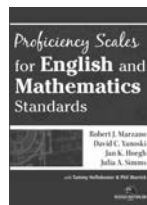
MRL2779



MRL2861



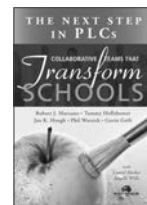
MRL1222



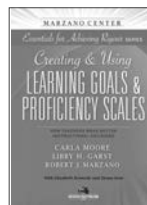
MRL2878



HB7477



MRL7484



LSM5626

Attention Order Number

Name of School

Address

..... State P/Code

Country

Email:

Yes, I would like to receive emails from Hawker Brownlow Education about future workshops, conferences and the latest publications.

Terms of Trade

- Prices are quoted in Australian dollars (\$AUD) and include GST
- All prices are subject to change without notice.
- For New Zealand customers, at the time of invoice, we will convert the amount into New Zealand dollars (\$NZD) so that you can pay by cheque or credit card in New Zealand dollars (\$NZD).
- Full money-back guarantee.
- We do realise it is difficult to order sight unseen. To assist you in your selection, please visit our website <www.hbe.com.au>. Go to 'Browse Books' and most titles will give you the option to view the first few pages of the book. Click 'View Contents' on your selected book page.
- We will supply our books on approval, and if they do not suit your requirements we will accept undamaged returns for full credit or refund. Posters are for firm sale only and will not be sent on approval. Please be aware that delivery and return postage is the responsibility of the customer.
- Freight costs are determined at Australia Post rates, with a minimum delivery charge of \$9.50 within Australia and \$15.00 for New Zealand for each order.
- Please provide your street address for delivery purposes.

To place an order, request a catalogue or find out more about our resources:

Call
1800 334 603
(03) 8558 2444

Fax
1800 150 445
(03) 8558 2400

Online
www.hbe.com.au

Mail
Hawker Brownlow Education
PO Box 580,
Moorabbin, VIC 3189

Do you want to know all about the latest professional development events in your area? Be the first to find out about new releases from world-renowned and local authors with the HBE e-newsletter! Upcoming titles will feature authentic assessment and digital media, along with a strong focus on success in mathematics and literacy. Sign up to our FREE e-newsletter at www.hbe.com.au.

Online 'On Account' ordering now available!

If you have a pre-existing account with Hawker Brownlow Education, you can now order online and pay using that account.