

Research Conference 2022

Reimagining
assessment

Proceedings and Program

ONLINE EVENT
22–25 August 2022



Program and Organising Committee

Dr Claire Scoular
Conference Convenor

Program committee

Professor Geoff Masters AO, CEO, ACER
Dr Dan Cloney
Dr Jarrod Hingston
Dr Sandra Knowles
Dr Dara Ramalingam
Dr Sarah Richardson
Dr Pina Tarricone

Organising committee

Daniela Buehrig
Kylie Burns
Matt Bongetti
Alex Harkness
Emma Noble
Dr Ralph Saubern
Toby Simmer
Ian Teo
Kelly Van Os
Kelly Williams

Australian Council for Educational Research
19 Prospect Hill Road
Camberwell VIC 3124
Phone: +61 3 9277 5555
www.acer.org

ISBN 978-1-74286-685-7
<https://doi.org/10.37517/978-1-74286-685-7>
Australian Council for Educational Research © 2022

All rights reserved. Except under the conditions described in the *Copyright Act 1968 of Australia* and subsequent amendments, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the written permission of the publisher.

Design by ACER Creative Services
Edited by Kylie Burns

ABN 19 004 398 145

Acknowledgement of Country

In the spirit of reconciliation, ACER acknowledges the Traditional Custodians of Country throughout Australia and their connections to land, sea and community. We pay our respect to their elders past and present and extend that respect to all Aboriginal and Torres Strait Islander peoples today. ACER acknowledges the Aboriginal and Torres Strait Islander people who continue to contribute to our work to improve learning, education and research.

CONTENTS

Foreword	iv
Professor Geoff Masters AO	
Keynote papers	
Reimagining the purpose of assessment	1
Professor Geoff Masters AO	
An innovative method for teachers to formatively assess writing online	5
Dr Sandy Heldsinger and Associate Professor Stephen Humphry	
Assessment moderation: Is it fit for purpose?	12
Associate Professor Lenore Adie	
Karmel Oration	
Making learning visible: Moving from nouns to verbs	17
Dr Diane DeBacker	
Conference papers	
ACER models of writing: Changing the assessment mindset	20
Juliette Mendelovits and Judy Nixon	
Reimagining classroom assessment and feedback to meet learner needs	25
Dr Fabienne Van der Kleij	
Using PAT data to inform teaching and learning	30
Marc Kralj, Rachel Felgate, Dr Shani Sniedze-Gregory, Caithlin Power, Grant Barclay and Darren Leech	
Sharing and securing learners' performance standards across schools	39
Emeritus Professor Richard Kimbell	
Assessing reading: How assessment can be used to target teaching and enhance understanding of reading comprehension	45
Dr Sandra Knowles	
Teachers' assessment literacy and design competence framework	50
Dr Sladana Krstic, Dr Sarah Manlove and Dr Sarah Richardson	
Conference program	57

FOREWORD

Reimagining assessment

Welcome to Research Conference 2022



ACER organises the annual Research Conference to review and discuss the latest research findings in a key area of education policy and practice. The focus of this year's Research Conference is on the use of assessment to support improved teaching and learning. The conference is titled 'reimagining assessment' because we believe there is a need to transform the essential purposes of educational assessment to provide better information about the deep conceptual learning, skills, competencies, and personal attributes that teachers and schools now have as objectives for student learning and development. Reimagined assessments must now be focused on monitoring learning across this broader range of intended outcomes and provide quality information about the points individuals have reached in their long-term development.

ACER has invited some leading educational researchers to join us to share the findings of their research relevant to this topic. This research includes investigations of how new technologies can be used to gather and analyse information about student learning and performance in real time to provide deeper understandings of the difficulties and misunderstandings students experience. It also includes research into the nature of increasing proficiency in newly prioritised areas of learning, and the design and implementation of innovative and promising new assessment methods to address these.

Our decision last year to move the Research Conference entirely online has enabled us to include a larger number of international speakers in our conferences and to profile a wider range of the latest research.

I trust that you gain significant professional benefit from this year's Research Conference.

A handwritten signature in black ink that reads "Geoff Masters". The signature is written in a cursive, flowing style.

Professor Geoff Masters AO

CEO, Australian Council for Educational Research

Reimagining the purpose of assessment

Professor Geoff Masters AO

Australian Council for Educational Research

<https://doi.org/10.37517/978-1-74286-685-7-10>

Professor Geoff Masters is Chief Executive Officer and a member of the Board of the Australian Council for Educational Research. His work with governments and school systems has included the development of strategies for improving literacy and numeracy learning; a review of upper secondary assessment and university entrance procedures; a major review of the New South Wales school curriculum; the development of a School Improvement Tool; and the development of a Principal Performance Improvement Tool. For three decades he has written extensively on the reform of educational assessment processes, including in Reforming educational assessment: Imperatives, principles and challenges. Recently he led an international study of five high-performing school systems for the US National Center on Education and the Economy. His contributions to education have been recognised through the award of the Australian College of Educators' Medal and his appointment as an Officer of the Order of Australia.

Abstract

Assessment is commonly understood as the process of judging how well students have learnt what they have been taught. It comes at the end of a sequence that begins with a curriculum or course syllabus. Teachers are expected to deliver this body of specified content, students are expected to learn it, and assessment is the process of judging and grading students on how well they have learnt what teachers have taught. This is a common view of assessment among students, parents and many teachers.

I will argue in this presentation that this traditional understanding of what it means to learn successfully is no longer serving us well. It results in many less advanced students falling further behind in their learning, being written off as poor learners, and eventually disengaging from school. It also fails to challenge and extend some of our most advanced students, resulting in less progress than they are capable of making.

The alternative is to use assessment to *monitor* learning – that is, to establish and understand where learners are in their long-term learning progress; to evaluate growth over time; and to provide teachers, students, parents, school leaders, system managers and governments with quality information to promote further learning. Shifting the focus from judging and grading to *monitoring* would be a significant change in assessment practice with the potential to better support successful learning.

Establish where students are in their learning

The monitoring of learning requires assessment processes that are designed to establish and understand where students are in their learning at the time of assessment. In classroom settings, this information is essential for at least three reasons. First, students in the same year of school can be at widely different points in their long-term learning and development. The most advanced 10 per cent of students in any year of school are typically five to six years ahead of the least advanced 10 per cent of students. As a consequence, students commence each school year with very different starting points, levels of readiness and learning needs. Effective teachers are sensitive to these differences and work to establish and understand exactly where individuals are in their long-term progress.

Second, and related, successful learning is unlikely when individuals are given material that is much too difficult or much too easy. Inappropriately difficult material leads to frustration and disengagement; inappropriately easy material leads to boredom and lack of effort. Successful learning is most likely when learners are given stretch challenges beyond their comfort zones, in what Vygotsky calls the 'zone of proximal development', where success may depend on scaffolding and support (Vygotsky, 1978). Effective teachers maximise the probability of successful learning by providing individuals with well-targeted learning opportunities.

Third, students often differ in their skill gaps and understandings of subject matter. It is now well established that learners create their own mental models of what they are learning. These models can be different from learner to learner, as well as being markedly different from the understandings that teachers assume. Student misconceptions often remain hidden and become obstacles to further learning progress. Effective teachers work to uncover and understand how individuals are thinking about subject matter and intervene to address specific errors and misconceptions.

Thus, effective teaching recognises that individuals can be at quite different points in their learning and may be progressing at different rates. It is also underpinned by a deep belief that every student can make further progress if they can be engaged, motivated and provided with appropriate learning challenges. Many teachers intuitively believe this, but developing the skills, practices and supporting structures to make it happen across all schools requires intentional, large-scale effort.

The process of establishing where learners are in their learning depends on a deep understanding of the relevant learning domain. Progress usually takes the form of increasing 'proficiency' reflected in more extensive knowledge, deeper understandings and higher-level skills. Establishing where students are in their learning at any given time means clarifying what they know, understand and can do at that point in their learning. This can be done at the level of an entire domain such as mathematics, for particular sub-domains such as number, space/geometry and algebra, or in relation to specific skills and understandings.

Most learning domains extend over many years of school. Some begin before, and continue beyond, school. Examples include the broad domains of reading, mathematics and science. Most general attributes and capabilities, such as the ability to work in teams, to communicate, and to create innovative solutions to problems, also develop over extended periods of time. Because students of the same age tend to be at widely varying stages in their learning progress, teachers require good understandings of the nature of long-term development if they are to ensure that every student is appropriately challenged and extended.

When students understand what higher levels of proficiency and better quality work look like in an area of learning, they have a basis for reflecting on their current levels of achievement and for setting realistic, challenging goals for further learning. In this way, students are supported to take a degree of responsibility for their own learning and long-term progress.

Evaluate progress over time

The monitoring of learning also depends on the ability to evaluate the *progress* students make over time. At a minimum, information is required at two time points; for example, before and after a course of instruction. Ideally, progress is monitored over multiple time points to reveal long-term learning trajectories.

In practice, attempts to evaluate learning sometimes are based on information collected at only one time point – for example, at the end of a period of instruction. However, unless all students commence a period of instruction at the same point in their learning (a highly unlikely scenario in actual school settings) information collected only at the end of instruction is inadequate for comparing and evaluating how much students have learnt.

Information about the progress an individual makes can be used to judge the adequacy of that student's learning. If a student is making no progress, or very little progress over time, that may indicate lack of effort or a particular difficulty that the student is experiencing (e.g. the absence of prerequisite skills or understandings). By studying rates of long-term progress, it is possible to identify students who are on learning trajectories that have plateaued and who are slipping behind in their learning. Such information is crucial for effective intervention.

Current approaches to organising and delivering schooling often deny students the opportunity to see and appreciate the long-term progress they are making. Schooling usually is divided into school years, semesters, courses and learning modules, which are often treated as discrete and unrelated. Students are judged and graded on the content taught in each of these learning periods, often with each new period being treated as a fresh start. As a result, less advanced students can receive low grades year after year, reinforcing low self-perceptions and providing little indication of the long-term progress they actually make. At the other extreme, more advanced students can be judged to be performing well on year-level expectations, and so receive high grades, but make relatively little year-on-year progress.

Perhaps the most effective way to build individuals' confidence in their ability to learn is to assist them to see the progress they make over time – possibly over multiple years of school. Long-term pictures of progress allow students to appreciate how the quality of their work has improved and how they can now perform tasks that once were beyond them. Information about improvement also gives students a better understanding of the relationship between effort and success.

Policy implications

Changing the focus of assessment from judging and grading to monitoring requires a series of processes.

Underpin curriculum and assessment processes with empirically-based learning progressions

Monitoring depends on an empirically-based learning progression that describes and illustrates the nature of long-term progress in an area of learning. A learning progression normally describes the development of increasingly sophisticated knowledge, deeper understandings, and higher levels of competence and skill (including critical thinking, creative thinking, problem-solving, and collaborating). It provides the basis for the sequencing of the curriculum and a frame of reference for monitoring student progress over time.

Develop tools to support teachers to establish the points individuals have reached in their learning, with related teaching resources

To understand where students are in their learning and to design interventions to promote further learning, teachers require quality tools and resources built for this purpose. They also require skills in using these tools.

Develop forms of reporting that indicate where students are in their long-term learning and that assist in monitoring learning progress over time

We can no longer afford to have large numbers of students written off as poor learners because they perform below year-level expectations year after year. The alternative would be to have reports that identify the points individuals have reached in their learning, indicate what might be done to support further learning, and provide information about learning progress over time. The expectation should be that every student will make excellent progress every year, regardless of their starting point.

Implement systems for recording, tracking and certifying student attainment independently of year levels and phases of schooling

It is likely that, as wider use is made of technologies to support learning, students will increasingly be able to learn at their own pace. As a result, there will be a need for better ways of tracking and recording the points individuals have reached in their learning. There will also be a need for new ways of recognising and certifying achievement. Students increasingly will be judged against performance standards that are not tied to particular years of school (or even phases of school). Just as music students are able to take assessments when they feel ready to demonstrate performance at the next level, students may move through levels of certification at their own pace. In general, the assessment systems that would enable this do not currently exist and need to be developed.

References

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

An innovative method for teachers to formatively assess writing online

Dr Sandy Heldsinger and Associate Professor Stephen Humphry

University of Western Australia

<https://doi.org/10.37517/978-1-74286-685-7-1>

Dr Sandy Heldsinger is leading the implementation of the Brightpath assessment and reporting software in schools. Sandy coordinated the Western Australia system-level assessments, has taught a masters-level course in educational assessment for a number of years and has led the development of a wide range of resources, including reporting software, to support schools in using assessment to improve student performance. Sandy recently developed central components of the Western Australian Curriculum Outline for WA's School Curriculum and Standards Authority and she coauthored What Teachers Need To Know About Assessment and Reporting, published by ACER.

Steve Humphry is an Associate Professor in Educational Assessment, Measurement and Evaluation at The University of Western Australia (UWA). He won a position with UWA in 2006 and has held substantial Australian Research Council Linkage grants continuously since 2008. He has worked with industry partners on these grants, including the Australian Curriculum Assessment and Reporting Authority, UNESCO's International Institute for Educational Planning, and the School Curriculum and Standards Authority. Stephen has extensive experience in the NAPLAN and WALNA (Western Australian Literacy and Numeracy Assessment) large-scale testing programs. His research has focused increasingly on developing novel approaches that allow classroom teachers to reliably assess students in areas not amenable to large-scale testing such as visual art and science investigations. In industry work, between 2011 and 2013, Stephen led the Central Analysis of Data project for NAPLAN, on which a team of psychometricians established NAPLAN literacy and numeracy scales that form the basis for reporting on performance at the Australian, state and territory, school and individual student levels.

Abstract

Assessment is an integral component of effective teaching and a teacher's professional judgement influences all routine aspects of their work. In the last 20 years, there has been considerable work internationally to support teachers in using assessment to improve student learning. However, there is a pressing issue that impedes teachers' professional judgement being exploited to its full potential. The issue relates to teacher assessment of learning progression in the context of extended performances such as essays and arises from the complexity of obtaining reliable or consistent teacher assessments of students' work. Literature published in the United States, England and Australia details evidence of low reliability and bias in teacher assessments. As a result, despite policymakers' willingness to consider making greater use of teachers' judgements in summative assessment, and thus provide for greater parity of esteem between teacher assessments and standardised testing, few gains have been made. Although low reliability of scoring is a pressing issue in contexts where the data are used for summative purposes, it is also an issue for formative assessment. Inaccurate assessment necessarily impedes the effectiveness of any follow-up activity, and hence the effectiveness of formative assessment. In this session, we share our research of writing assessment and explain how it has led to the development of an innovative assessment process that provides the advantages of rubrics, comparative judgements, and automated marking with few of the disadvantages.

Introduction

Despite the widespread desire for teacher judgements to be used for summative assessments, attaining reliable judgements has been a challenge (Brookhart, 2013; Harlen, 2004; Johnson, 2013). Instead, external standardised assessments are mostly used for this purpose. Similarly, although assessment is an integral component of teaching, and professional judgement influences various aspects of teachers' work (Black & Wiliam, 2010; Du Four, 2007; Hattie & Timperley, 2007), the reliability of formative assessments is seldom examined.

A growing body of research shows that teachers make reliable judgements by making pairwise comparisons of extended performances (Humphry & Heldsinger, 2019). Using this approach, teachers compare pairs of performances and judge which performance, in each pair, demonstrates a higher level of attainment. Performances are placed on the scale from weakest to strongest, empirically showing learning progression. The terms comparative judgement, comparative pairs, and paired comparison are also used to describe pairwise comparisons (Tarricone & Newhouse, 2016).

A drawback of pairwise comparisons is that they are time-consuming as a method for teacher assessment (Bramley et al., 1998). In addition, pairwise comparisons provide the basis for scaling and ordering of student performances but the approach does not directly avail teachers of diagnostic information in a form that can be acted upon. However, once performances have been ordered, they can be qualitatively examined and doing so provides insight into changes in features of writing observed with increasing development. Thus, the application of pairwise comparisons potentially provides the basis both for internally consistent judgements and diagnostic information. The method described in this article is designed to provide these advantages to classroom teachers. As described to follow, the two-staged method is designed so that it is time-effective, accessible, and provides immediate and actionable formative feedback.

The two-stage assessment method

Constructing scales using pairwise comparisons

The first stage in the two-stage assessment method is to calibrate a scale and subsequently to select exemplars. The literature provides background on the use of the method of pairwise comparisons in education and other fields (Bond & Fox, 2001; Bramley et al., 1998; Pollitt, 2012; Thurstone, 1927, 1959). In Stage 1, assessment tasks are administered by classroom teachers and a relatively large number of performances are collected. Teachers compare these performances and select which performance is on-balance better given key performance features to be considered. The pairwise comparison data are analysed using the Bradley-Terry-Luce (BTL) model (Bradley & Terry, 1952; Luce, 1959) to produce a performance scale. Scale locations are inferred from the proportions of judgements in favour of each performance when compared with others. If all performances were compared with each other, the strongest performance would be the one judged better than the other performances the greatest number of times.

However, in practice, scaling techniques can be used and it is unnecessary for each performance to be compared with every other performance. Data from pairwise comparisons are analysed to examine the overall internal consistency. Data are also analysed to ascertain whether each teacher's comparisons are consistent with the overall scale locations, within expectations given the BTL model. Specifically, the Person Separation Index is used to examine internal consistency and fit indices are used to examine the consistency of teachers' comparisons with overall scale locations, as reported, for example, in Humphry and Heldsinger (2020).

Fit to the BTL model is also examined for each performance on the scale. Fit indices and qualitative examination of performances are used to select a set of exemplars for use in Stage 2, in which teachers assess other performances against the scale. Performances are not used as exemplars if the pairwise comparisons produce data that departs too much from the Guttman pattern (Guttman, 1944). Performances with relatively Guttman-like patterns are compared consistently with overall ordering and provide better exemplars for Stage 2. These performances are more clearly ordered and provide a clearer reference point for assessment against the scale.

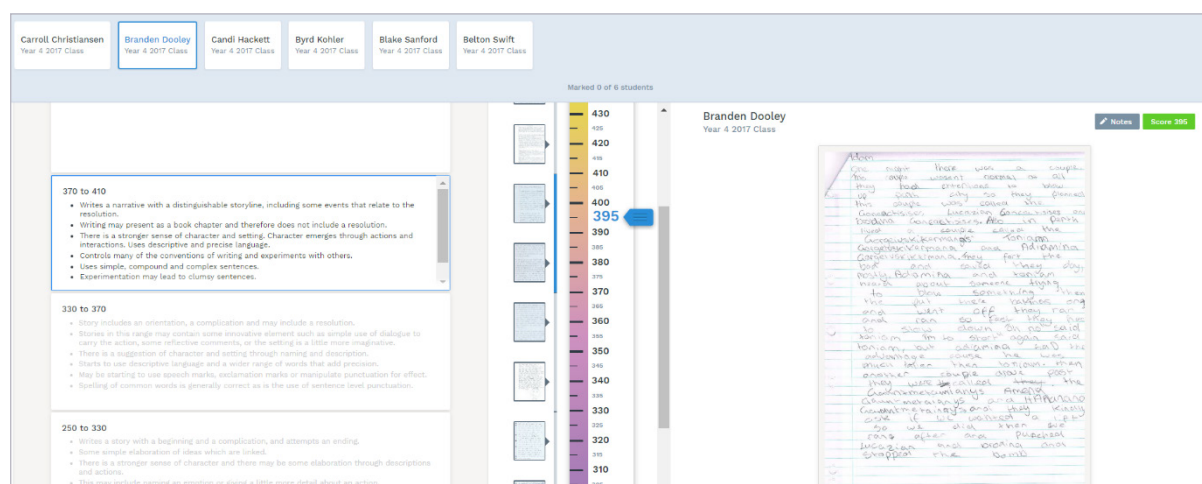
Exemplars and descriptors

Once performances have been placed on a scale, they are placed in order of location on the scale physically (e.g. on a table) and examined to infer the features and levels of writing that students in a given range of the scale demonstrate in their performances. In this way, learning progressions are described. In contrast with typical rubrics, performance descriptors are based on a systematic analysis of the performances placed in order according to one or more criteria. This enables teachers to glean empirically based information about how performances change with progression from lower to higher levels. It also provides a basis for specific feedback on key points, which are referred to as Teaching Points. Descriptors focus on features that are most relevant in a given range of the scale.

The teacher's ruler is an interactive display comprising several key elements, as shown in Figure 1. Teachers assess their own students' performances, shown on the right-hand-side, against the ordered exemplars, which are shown in the centre of the screen. Teachers refer to the empirically based descriptors displayed on the left-hand side. Teachers can click on exemplars to expand and view them on the left-hand side of the display. To assess their students' work, teachers locate where a performance is likely to sit on the scale based on comparisons with the exemplars and using the descriptors as a guide.

The interactive display provides the advantages of rubrics and comparative judgements. Specifically, teachers refer to general descriptions of performances relating to a given range, and they also compare performances with real, pre-calibrated exemplars.

Figure 1 Teacher's Ruler display



Teachers make an on-balance judgement based on their analysis of the strengths and weaknesses of the performance, and to determine which exemplar the performance was closest to or which two exemplars it fell between. In the display shown in Figure 1, exemplars are displayed in order from lowest to highest. Teacher comparisons are implicit rather than explicit. For example, if a performance is judged above the 10th exemplar but below the 11th exemplar in the Teacher's Ruler, it is implied that the performance is better than all exemplars below the 10th and worse than all exemplars above the 11th. The performance is given the scale score associated with or above or below the exemplar. The scale is shown in the centre of Figure 1. The scale locations are based on the analysis of the pairwise comparison data; specifically, they are transformations of the logit scale obtained from analysis of data using the BTL model.

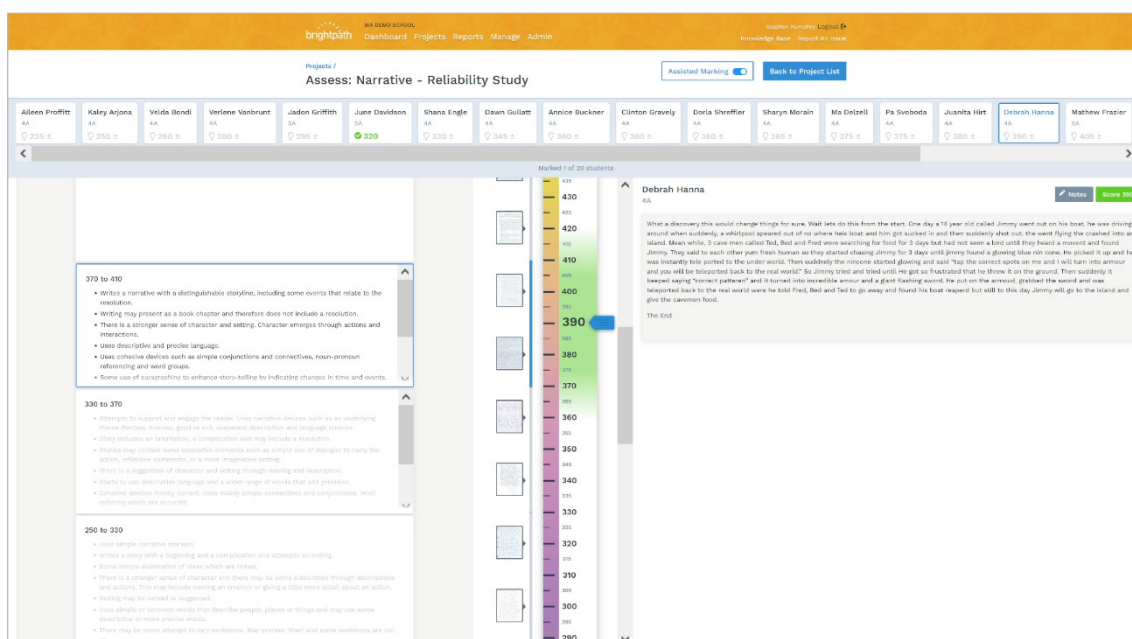
Judges are provided with a guide to help make their judgements. This guide contains all the calibrated exemplars, the performance descriptors, and a close qualitative analysis of each exemplar.

Assisted marking with Natural Language Processing and calibrated exemplars

Automated scoring is often used instead of human marking or to check human marking. However, it is not necessary to adopt a process in which human and automated scoring occur separately. An automated Marking Assistant has been designed to help teachers quickly focus in the right zone of the Teacher's Ruler in much the same way that a search-suggestion helps users focus on information that is most relevant. This process is designed to help teachers to concentrate on features of writing that are best judged by humans.

Based on Natural Language Processing (NLP) indices, the automated assistant predicts scale locations of online performances as a starting point. In Figure 2, the blue arrow pointer shows the predicted score for the performance on the right. The green zone shows the interval or range in which the performance's score is predicted to lie with approximately 70 per cent confidence. For teachers who are not yet familiar with the exemplars, the predictions enable efficient assessments from the start, while teachers gain familiarity. For teachers who are familiar, the predictions serve as a reference point. Teachers can also turn the predictions off.

Figure 2 Marking Assistant prediction on the display



Moderation

The Teacher's Ruler uses exemplars to show what constitutes a given score. Teachers in one school or classroom in a school see the same set of ordered exemplars as teachers in another school or classroom. If the assessments are conducted effectively, they are automatically moderated. Various issues may occur with rubrics depending on how they are designed and constructed (Humphry & Heldsinger, 2014). Bias and rating tendencies that are common in rubrics are limited by having exemplars as the basis for implicit pairwise comparisons. In rubrics, performances are referenced to descriptions, and the descriptions can be interpreted differently by different teachers. In pairwise comparisons, it is difficult to introduce bias because one performance is directly compared with another.

Formative assessment

Consistent with the logic of the cumulative ordering in the Guttman pattern (Guttman, 1944), the following approach is used. For students in a given range of the Teacher's Ruler scale, descriptors applicable to students in the next higher range are used as teaching points. The rationale is that descriptors in the next higher range are most likely to describe what lies in a student's zone of proximal development with learning progression.

Teachers can refer to specific features of exemplars in providing feedback to students. Descriptors convey a general sense of performance; whereas the exemplars show, in more tangible and specific terms, what performances at a given level look like. The exemplars explicitly show different levels of performances in a way that is difficult to fully capture using descriptions of the kind that appear in rubrics. Together, descriptors and exemplars convey learning progression better than either individually.

Reliability of teacher judgements

Several studies have been conducted to examine the reliability of teacher judgements of narrative, persuasive, and information-report writing assessment using the second stage of the two-stage method. In each of these studies, all participants assessed a common set of approximately 25 performances using the Teacher's Ruler. Each marker's scale scores for the common performances were compared with the average scale scores given by all the other markers in the study. The correlations obtained from these studies are shown in Table 1 and show high levels of reliability using the Teacher's Ruler to assess the extended performances.

Table 1 Summary of reliability of the second stage assessments in a number of studies

	Study 1	Study 2	Other studies
Narrative	0.903*	0.927*	0.938
	n=12	n=37	n=65
Persuasive	0.848*	0.925	
	n=8	n=30	
Information report	0.966		
	n=34		

*previously published results

The evidence reported here is collected without use of Assisted Marking. Some of these results have been reported in published literature (Heldsinger & Humphry, 2013; Humphry et al., 2017; Humphry & Heldsinger, 2019; Humphry & Heldsinger, 2020).

Discussion

One requirement that has been present since the very first version, over a half century ago, is that tests should be adequately documented, the procedures by which tests were developed should be documented, and evidence regarding the validity of the tests, and specifically the reliability, must be produced (Black & Wiliam, 2012, p. 252).

Discussion of reliability in the context of teachers' assessments is often referred to as inter-rater reliability and relates to the generalisability of scores across markers or scorers. Differences that arise in scores that are not a function of student ability, but from differences in examiners, constitute a source of measurement error that negatively impact the reliability of the assessment. The key implication is that where differences in scores more accurately represent the differences in the construct being assessed, people can have more trust in scores when drawing inferences about students' ability and making decisions about follow-up actions.

The results outlined in this article provide empirical evidence that the two-stage method enables reliable teacher assessments, responding to calls for research into the reliability of teacher assessments by Harlen (2005), Brookhart (2013), and Johnson (2013). A negative impact of giving high-profile to external and standardised test-based results can be a loss of assessment skill on the part of teachers as well as loss of confidence in their ability to make sound assessments of their students (Black et al., 2010, 2011). A significant reason for placing emphasis on external and standardised assessments is the belief that teacher assessments are not sufficiently reliable. The key benefit of enabling teachers to make reliable assessments is that the professionalism of teachers is valued and fostered, consistent with the general desire to value teacher judgements observed by Johnson (2013).

In conclusion, the two-stage method of assessment enables teachers to make reliable judgements of writing. An advantage of the method over pairwise comparisons alone is that once a scale has been constructed, the average time to assess a performance is reasonably modest. The use of assisted marking further reduces assessment time by enabling teachers to focus on what they are best placed to assess in performances. Unlike external testing programs, by using the two-stage method classroom teachers assess their own students and provide formative feedback based on their own assessments and familiarity with the students' work.

References

- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2010). Validity in teachers' summative assessments. *Assessment in Education: Principles, Policy & Practice*, 17, 217-234. <https://doi.org/10.1080/09695941003696016>
- Black, P., Harrison, C., Hodgen, J., Marshall, B., & Serret, N. (2011). Can teachers' summative assessments produce dependable results and also enhance classroom learning? *Assessment in Education: Principles, Policy & Practice*, 18, 451-469. <https://doi.org/10.1080/0969594X.2011.557020>
- Black, P., & Wiliam, D. (2010). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 92(1), 81-90. <https://doi.org/10.1177/003172171009200119>
- Black, P., & Wiliam, D. (2012). The reliability of assessments. In J. Gardner (Ed.), *Assessment and learning* (2nd Edn, pp. 243-263). Sage Publications Ltd.
- Bond, T. G., & Fox, C. M. (Eds.) (2001). *Applying the Rasch model: Fundamental measurement in the human sciences*. Lawrence Erlbaum Associates Inc. <https://doi.org/10.4324/9781410600127>

- Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs, I. The method of paired comparisons. *Biometrika*, 39(3/4), 324–345.
- Bramley, T., Bell, J. F., & Pollitt, A. (1998). Assessing changes in standards over time using Thurstone's paired comparisons. *Education Research and Perspectives*, 25(2), 1–24.
- Brookhart, S. M. (2013). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 20(1), 69–90. <https://doi.org/10.1080/0969594X.2012.703170>
- Du Four, R. (2007). Once upon a time: A tale of excellence in assessment. In D. Reeves (Ed.), *Ahead of the curve. The power of assessment to transform teaching and learning* (pp. 253–267). Solution Tree Press.
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9(2), 139–150.
- Harlen, W. (2004). A systematic review of the evidence of reliability and validity of assessment by teachers used for summative purposes. In *Research evidence in education library*. EPPI-Centre, Social Science Research Unit, Institute of Education, University of London.
- Harlen, W. (2005). Trusting teachers' judgement: Research evidence of the reliability and validity of teachers' assessment used for summative purposes. *Research Papers in Education*, 20(3), 245–270. <https://doi.org/10.1080/02671520500193744>
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Heldsinger, S., & Humphry, S. (2013). Using calibrated exemplars in the teacher-assessment of writing: an empirical study. *Educational Research*, 55(3), 219–235. <https://doi.org/10.1080/00131881.2013.825159>
- Humphry, S., & Heldsinger, S. (2014). Common structural design features of rubrics may represent a threat to validity. *Educational Researcher*, 43(5), 253–263.
- Humphry, S., & Heldsinger, S. (2019). A two-stage method for classroom assessments of essay writing. *Journal of Educational Measurement*, 56(3), 505–520. <https://doi.org/10.1111/jedm.12223>
- Humphry, S., & Heldsinger, S. (2020). A two-stage method for obtaining reliable teacher assessments of writing. *Frontiers in Education*, 5. <https://doi.org/10.3389/educ.2020.00006>
- Humphry, S., Heldsinger, S., & Dawkins, S. (2017). A two-stage assessment method for assessing oral language in early childhood. *Australian Journal of Education*, 61(2), 124–140. <https://doi.org/10.1177/0004944117712777>
- Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education*, 28(1), 91–105. <https://doi.org/10.1080/02671522.2012.754229>
- Luce, R. D. (1959). *Individual choice behaviours: A theoretical analysis*. J.Wiley.
- Pollitt, A. (2012). The method of adaptive comparative judgement. *Assessment in Education: Principles, Policy & Practice*, 19(3), 281–300. <https://doi.org/10.1080/0969594X.2012.665354>
- Tarricone, P., & Newhouse, C. P. (2016). Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability. *International Journal of Educational Technology in Higher Education*, 13(1), 16. <https://doi.org/10.1186/s41239-016-0018-x>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34(4), 273–286. <https://doi.org/10.1037/h0070288>
- Thurstone, L. L. (1959). *The measurement of values*. The University of Chicago Press.

Assessment moderation: Is it fit for purpose?

Associate Professor Lenore Adie

Australian Catholic University

<https://doi.org/10.37517/978-1-74286-685-7-2>

Lenore Adie is Associate Professor of Teacher Education and Assessment, and Senior Research Fellow at the Institute for Learning Sciences and Teacher Education, Australian Catholic University. Her research focuses on assessment and moderation processes, which contribute to quality assurance and improvement purposes. Lenore's research has generated new knowledge in the field of assessment, focusing on quality in assessment practices and processes, especially within systems of standards-referenced assessment. She currently leads an Australian Research Council project investigating the application of standards in promoting learner growth. Lenore has over 30 years of professional experience working in schools as a teacher and in leadership positions, and in teacher education.

Abstract

This presentation outlines the different practices and purposes of assessment moderation and analyses the benefits and issues of these. Two projects are presented; each seeks to improve teacher judgement. Both attempt to redefine what we mean by moderation and how moderation can be conducted in ways to reach diverse and dispersed groups. The first draws on a current Australian Research Council project that is exploring the development of scaled exemplars, the use of commentaries of judgement decisions, and the use of digital platforms to support teachers grading student work. The second draws on the work of the Graduate Teacher Performance Assessment (GTPA) group of 19 universities to demonstrate how large-scale moderation of complex performance assessments can be conducted across states and territories. The layers of processes, including the use of technology, required to ensure judgement reliability are illustrated. In both contexts, the ways in which moderation is understood and practised are being reimaged.

Introduction

Assessment has long been a contentious practice. What to test, when and whose role it is, are embedded in levels of trust and beliefs in teacher professionalism. Moderation is one of the quality assurance processes we look to for reliability and consistency in judgement decisions. But this stance assumes that aspects such as consistency in judgement are desirable. Some teachers would view these practices as a form of top-down accountability that constrain their professionalism by imposing a structure that hinders their decision-making about how to best progress a student's learning. The position taken in this paper, however, is that assessment and moderation can be designed and used in ways that enhance the agency of teachers to improve their teaching and the learning of their students. Of importance here is how these practices contribute to, rather than distract from, good teaching that progresses student learning.

What is understood as assessment moderation?

Moderation, as a process, is broadly understood as a form of quality assurance and an 'approach to agreeing, assuring and checking standards' (Bloxham et al., 2016, p. 638). Moderation of judgement decisions can be undertaken in various ways, dependent on purpose. For example, consensus or social moderation is used to reach agreement on grades awarded. Expert moderation defers to the evaluative experience of someone such as a curriculum leader to make a final judgement decision, particularly in cases of disagreement about the awarded grade. Statistical moderation, often used in high-stakes examinations, has a focus on reliability and involves the scaling of scored performances (Crisp, 2017).

Social moderation meetings involve teachers in discussion and negotiation of the meaning of criteria and standards to reach agreement of an awarded grade. Through these discussions, teachers can develop shared understandings of the qualities of a standard and the different ways these can combine while still representing the same standard. This process of discussion and negotiation is essential since standards are text-based, so various interpretations of terms can exist. Individual histories impact on the reading of any text, and research has shown that teachers have different understandings of standards (Wyatt-Smith & Gunn, 2009). Through social moderation, teachers can reach agreement on the qualities that represent a standard. The issue is that these shared understandings of an expected standard of performance can remain idiosyncratic to a school, or to a district, and in fact, vary widely across schools and districts.

Statistical moderation is claimed to mitigate the subjectivity of human judgement by calibrating scores against other assessments or judgement patterns. In this way, comparability across dispersed populations can be achieved. However, a recent example illustrates the challenges that are made to the claim of objectivity. During the COVID-19 pandemic, the cancellation of end-of-course examinations in the United Kingdom turned the gaze to teacher knowledge of student performance. A decision was made to use teachers' estimation of grades based on student classroom performance with an algorithm to adjust these scores based on the historical examination performance of the school (Kippin & Cairney, 2022; Porter, 2020). In a turn of events, the algorithm was charged with introducing bias as it caused the socio-economic inequalities among schools to be perpetuated through its use of historical performance data. This example illustrates how statistical models may not, of themselves, remove bias, and that data should always be interrogated for what it can tell us.

How can moderation processes be strengthened to meet different purposes and provide reliable outcomes?

In our research, we have been investigating how statistical and social moderation can be utilised to improve consistency and demonstrate reliability. To reach across dispersed populations, we have used and developed customised online infrastructures. To address diversity within these populations, several key practices and artefacts have been identified. We present two current projects to illustrate how moderation can be implemented in ways that connect teacher judgement with improved teaching and learning. Both have involved us in rethinking our understanding of moderation.

Project 1: Improving teacher assessment capabilities using exemplars and online moderation¹

This project, funded by the Australian Research Council, is investigating the extent to which the development of scaled exemplars of A–E standards with commentaries of how the judgement was made supports comparability of judgement. The research focuses on the middle years of schooling, specifically Years 4, 6 and 8, in the areas of English narrative, maths and science investigations and religious education.

The project involves teachers working online in:

- a process of pairwise comparison of submitted samples of student assessments.
- a standard-setting process to identify samples that best represent a standard descriptor.
- writing cognitive commentaries for selected exemplars. The commentaries describe the strengths and weaknesses of a work sample and how these combine to give an overall judgement, as well as identifying next steps for teaching (Wyatt-Smith & Adie, 2021). Teachers have told us that writing the commentaries focused their attention on specific features of assessments and the application to next steps for teaching. This has given them the knowledge to progress learning from one assessment to the next unit of work. Commentaries also provide a means for this thinking to be used by others.
- meeting online with other year-level and discipline teachers to discuss and refine the cognitive commentaries.
- using the exemplars with the accompanying cognitive commentary to moderate their own judgement decisions. A selected set of A–E student work samples will be blind reviewed by discipline experts to further evaluate the usefulness of the resources to support consistency in teacher judgement.

Across the project, various quantitative and qualitative methodologies have been used to identify consistency of teacher judgements and identify exemplar samples. Forms of statistical and social moderation have occurred across the different activities. The proposed outcome is a form of online moderation that takes forward teachers' explanations of their judgement decisions through the cognitive commentary to enable teachers across diverse locations to moderate their own judgement decisions.

Project 2: The Graduate Teacher Performance Assessment (GTPA®)²

Since 2019, Australian initial teacher education (ITE) providers have been required to implement a teaching performance assessment to final-year preservice teachers, with successful completion required for graduation and licensure. The GTPA is one such validated assessment, led by the Australian Catholic University, and used by 19 universities across Australia. Supporting consistency of judgement across a large and diverse group of universities involved the design of customised digital infrastructure and app design (Wyatt-Smith et al., 2022).

1 ARC LP180100046. We acknowledge the funding received for this project from the Australian Research Council, Australian Catholic University, University of Western Australia and Partner Organisations: the Queensland Department of Education, Western Australian School Curriculum and Standards Authority, and Catholic Education Western Australia Limited. The research team includes Associate Professor Lenore Adie, Professor Claire Wyatt-Smith, and Professor Michele Haynes (Australian Catholic University), Associate Professor Stephen Humphry and Professor David Andrich (University of Western Australia), Professor Chris DeLuca (Queen's University, Canada), and Mr Terry Gallagher (Queensland Department of Education).

2 The Graduate Teacher Performance Assessment (GTPA®) was created by the Institute for Learning Sciences and Teacher Education (ILSTE), Australian Catholic University and has been implemented in a Collective of Higher Education Institutions in Australia (graduatetpa.com). ILSTE has led the validation of the instrument, standard-setting and cross-institutional moderation with the engagement of teacher educators, policy personnel and a multidisciplinary research team.

At the centre of these processes is online cross-institutional moderation (CIM) that is conducted annually (CIM-Online™)³. CIM involves teacher educators from each university within the GTPA Collective individually scoring online de-identified samples from each university within the group (Wyatt-Smith & Adie, 2021). To ensure reliability, supporting processes and artefacts are provided, including:

- performance-level descriptors that support teacher educators to gain a sense of overall level, and criteria specifications that focus the gaze on required features
- calibration processes where judges grade previously validated exemplars prior to undertaking CIM to gain a sense of the expected standard at three levels of performance (meets, above and below the standard).
- exemplars in the form of validated preservice teacher samples from a previous moderation process with accompanying cognitive commentaries of judgement decisions that illustrate application of the standard used in calibration and for review when making judgements.
- anchor samples used to track any change in the applied standard over time.

The data from CIM are analysed statistically to investigate the comparability of judgements in terms of applying the established standard and endorsing submitted samples, as well as the performance characteristics of preservice teachers across program characteristics. These analyses are provided back to universities through a confidential report to use in ongoing programming decisions and for consideration in their teaching (formative purpose). The data are also used for accreditation purposes by the universities when reporting to state regulatory authorities (summative purpose).

For the first time, cumulative data from CIM-Online have provided information about the quality of teacher education graduates. Our data collected from 2017 have shown that the endorsement of samples increases over time though this is dependent on factors such as staff changes and university funding models. The collated data have contributed to rigorous professional conversations among teacher educators of the expected qualities for graduating preservice teachers, as well as investigation of context-specific responses within each university.

Conclusion

So, what have we learnt? By combining statistical and social processes of moderation with digital technologies and customised supporting resources, the consistency and reliability of judgement decisions can be improved over time. Significantly, through these processes, teachers have moved beyond moderation as an end process following summative assessment to one that views moderation as an ongoing process throughout teaching and learning, informed by the use of data for formative, improvement purposes. The main purpose of moderation may just be the opportunities it affords to interrogate and improve teaching and learning.

³ The online model of cross-institutional moderation (CIM-Online™) has been supported by digital architects in the Institute for Learning Sciences and Teacher Education, Australian Catholic University.

References

- Bloxham, S., Hughes, C., & Adie, L. (2016). What's the point of moderation? A discussion of the purposes achieved through contemporary moderation practices. *Assessment and Evaluation in Higher Education*, 41(4), 638–653. <https://doi.org/10.1080/02602938.2015.1039932>
- Crisp, V. (2017). The judgement processes involved in the moderation of teacher-assessed projects. *Oxford Review of Education*, 43(1), 19–37. <https://doi.org/10.1080/03054985.2016.1232245>
- Kippin, S., & Cairney, P. (2022). The COVID-19 exams fiasco across the UK: Four nations and two windows of opportunity. *British Politics*, 17(1), 1–23. <https://doi.org/10.1057/s41293-021-00162-y>
- Porter, J. (2020, 17 August). UK ditches exam results generated by biased algorithm after student protests. *The Verge*. <https://www.theverge.com/2020/8/17/21372045/uk-a-level-results-algorithm-biased-coronavirus-covid-19-pandemic-university-applications>
- Wyatt-Smith, C., & Adie, L. (2021). Introducing a new model for online cross-institutional moderation. In C. Wyatt-Smith, L. Adie, & J. Nuttal. (Eds). *Teaching performance assessments as a cultural disruptor in initial teacher education: Standards, evidence and collaboration* (pp. 39–58). Springer.
- Wyatt-Smith, C., Adie, L., Haynes, M., & Day, C. (2022). *Professionalizing teacher education: Performance assessment, standards, moderation, and evidence*. Routledge.
- Wyatt-Smith, C., & Gunn, S. (2009). Towards theorising assessment as critical inquiry. In C. Wyatt-Smith & J. J. Cumming (Eds.), *Educational assessment in the 21st century: Connecting theory and practice* (pp. 83–102). Springer.

Karmel Oration

Making learning visible: Moving from nouns to verbs

Dr Diane DeBacker

University of Kansas, United States

<https://doi.org/10.37517/978-1-74286-685-7-3>

Dr Diane DeBacker is the founding director of the Center for Certification and Competency Based Education (C3Be), an education research centre within the Achievement & Assessment Institute at the University of Kansas. C3Be focuses on responsive learning systems including learning maps, stackable micro-credentials, credit for prior learning, and authentic, outcomes-based assessments. Dr DeBacker's passion for education has taken her from her first high school classroom in Manhattan, Kansas, to advising the Abu Dhabi education system in the United Arab Emirates, from Chief Academic Officer in Seattle Public Schools to serving as Commissioner of Education for the State of Kansas, and as Executive Director of Business and Education Innovation for the Kansas Department of Commerce. In her 40 years in education, DeBacker has provided clarity to complex education issues including state education governance structures, school finance, federal education legislation, education policy, and workforce development and training.

Abstract

This presentation focuses on the need to make learning visible for all constituents: learners, educators, employers, policymakers, and future employees. As K–12 and higher education move to a post-pandemic world, it's imperative we find different ways to identify learner outcomes and evaluate learning. By moving from nouns to verbs; by moving from course titles to competencies; by moving from transcripts to learner outcomes, learning becomes visible. This presentation will focus on the work of the Center for Certification and Competency-Based Education (C3Be) at the University of Kansas in the United States. It discusses how learning is made visible through a proven process of mapping learning outcomes, assessing both new and prior learning using innovative technology, issuing stackable micro-credentials, and working with policymakers and employers to meet workforce demands.

Discussion

Course titles. Grade Levels. Grades. Transcripts. Education across all levels has historically been represented in examples such as these. Algebra 1. First Grade. Literature 101. A-B-C-D-F. Excellent. Satisfactory. Unsatisfactory. While titles and achievement reporting systems like these are familiar to students, parents, school personnel, and the public in general, they tell only a small portion of a learner's journey – the beginning (course titles and grade levels) and the end (grades and transcripts). As the world discovered during the COVID-19 pandemic, teaching and learning became more visible with the use of remote learning. What was once done in school buildings and, in many cases, behind closed classroom doors, became public. Parents, guardians, siblings, and anyone within a household could hear and see what educators were teaching and how learners were responding.

This deprivatisation of education certainly made teaching and learning more visible. Course titles and grade levels became blurred, especially for households with multiple children learning at the same time and in the same environment. Remote learning made even more obvious than before the necessity for education systems to better address the needs of learners and educators.

There are many components of a system to better address the needs of learners and educators – ideas such as competency-based education, stackable micro-credentials, just-in-time learning, demonstration of learning through authentic assessments, personalised learning, and credit for prior learning through life experiences are a few examples. C3Be has been exploring some of these ideas through proof-of-concepts. Specifically, a proof-of-concept learning map in cyber defense, a proof-of-concept process for issuing stackable micro-credentials, and a proof-of-concept for issuing credit for prior learning.

Learning maps

Learning maps are an effective tool for making learning visible as they go beyond a course title to the competencies (knowledge, skills, abilities) required for successful completion of a course. Learning maps are created from a synthesis of research within the domain or discipline, a structure that defines that domain, a set of hypotheses, and a mechanism for supporting growth. (DeBacker & Dudek, 2021). Learning maps include nodes (the competencies) and connections (the relationship between nodes). Together nodes and connections create clusters and neighborhoods of related knowledge and skills. In the proof-of-concept map completed by C3Be for an undergraduate cyber defense course, learning was made visible by moving from four course outcomes to approximately 140 nodes and connections. Educators, learners, and employers have all benefited from this more detailed view of a course within the profession of cybersecurity.

Micro-credentials

Micro-credentials are another way to make learning visible. Micro-credentials are related to a formally approved or accepted set of standards or competencies. They can be ‘stacked’ up to achieve a credential, certification, or degree that is recognised both within education systems, and business and industry. As such, micro-credentials offer a more granular way to move through the content/competencies than a traditional degree or certification program. Micro-credentials are also a way to address equity. The Credential As You Go initiative at SUNY State College (Saratoga Springs, New York, United States) aims to attract students who have some higher education but not a degree. From the SUNY State College website, ‘Learning not sealed within a credential is often not valued and remains as undocumented learning, reducing employment opportunities especially for minority and low-income students’.¹ Micro-credentials offer a timelier and more formative unit from which to be assessed resulting in lower stakes and lower cost for the learner. (DeBacker et al., 2021).¹ The process for issuing micro-credentials includes a thorough examination of the entire course of study leading to a certificate, credential, degree, or license. Included in this examination is discussion of the appropriate chunks of academic content to be considered and how many smaller chunks of information comprise a micro-credential. Finally, a decision on who will issue the micro-credential – an educational system, a professional organisation, a credentialing company – must be made. The micro-credential proof-of-concept work done by C3Be took place over the course of eight months as we collaborated weekly with the staff from Dynamynd (Scottsdale, Arizona, United States) to examine a youth specialist advocate professional development program in conation. The project included the review of existing course materials, the development of learner outcomes, revision of assessments, and options for issuing micro-credentials.

¹ Credential As You Go Symposium | SUNY Empire State College (esc.edu)

Credit for prior learning

Despite the many initiatives by education to acknowledge and address prior learning, we tend to return to what we are most familiar with: a traditional education system that starts all learners at the same place, keeps them on the same track, and expects them to reach the destination at the same time. The willingness of formal education programs to acknowledge an individual's prior learning regardless of where and how the learning took place has not been extensively or widely explored until recently. The COVID-19 interruption of learning resulted in learners, specifically learners in higher education, to re-examine their educational pursuits. As they experienced during the pandemic, learning was accessible 24 hours a day and took place beyond classrooms and campuses. While distance/remote/online learning has been around for decades, it has not been universally available, especially to underrepresented learners. Additionally, as unemployment rates skyrocketed during that time, employers began to consider qualifications other than a traditional college degree more seriously. Credit for prior learning (CPL) can be achieved by a variety of avenues including credit for military service; credit by standardised exams such as Advanced Placement, International Baccalaureate, or the College Level Examination Program; credit for industry credentials, certifications, or licenses; credit for portfolio(s); and credit by exams such as end-of-course exams. In a proof-of-concept CPL project for a bachelor's degree in information technology at the University of Kansas-Edwards campus (Overland Park, Kansas), C3Be has partnered with faculty and staff on an alignment study of industry-recognised certifications to courses within the degree program. It is anticipated that up to nine credit hours will be available for CPL beginning the fall semester of 2022.

Through the use of learning maps, stackable micro-credentials, a system for acknowledging prior learning and lived experiences, innovative technology, and partnering with employers and others, learning and the learner's journey is made visible.

References

- Credential As You Go Symposium | SUNY Empire State College (esc.edu) retrieved May 14, 2022.
- DeBacker, D. & Dudek, J. (2021, November 4). *Making learning visible: Developing competency-based learning maps*. [Paper presentation]. 2021 AECT International Convention, Chicago, Illinois.
- DeBacker, D., Dudek, J., Patelis, T., & Kingston, N. (2021, June 22). *Evaluating learning for the multiple constituencies of higher education: A call for action, a call for research* [Paper presentation]. 7th International Conference on Higher Education Advances, Valencia, Spain. 13621-35996-6-PB.pdf

ACER models of writing: Changing the assessment mindset

Juliette Mendelovits and Judy Nixon

Australian Council for Educational Research

<https://doi.org/10.37517/978-1-74286-685-7-4>

Juliette Mendelovits is the Research Director of the Assessment and Reporting research program at ACER. Since she joined ACER in 1991, Juliette has been involved in the conceptualisation, implementation and management of large-scale assessment projects, specialising in literacy, the humanities and social sciences. She played a leading role in the development of the OECD's Programme for International Student Assessment reading domain, from its first cycle in 2000 until 2012. Juliette has conceptualised and led the development of a number of reading and writing assessment projects at ACER, including the writing frameworks for ACER's International Schools' Assessment, Monitoring Trends in Educational Growth and the Southeast Asia Primary Learning Metrics. From 2015 to 2018, Juliette was based in London as Research Director and General Manager of ACER UK.

Judy Nixon is the Assistant Research Director of the Assessment and Reporting (Humanities and Social Sciences) research program at ACER. Since commencing at ACER in 2003, Judy has worked on the International Schools' Assessment, the Victorian General Achievement Test, the ACT Scaling Test, NAPLAN, NAP Civics and Citizenship, and a range of high-stakes selective tests and tertiary medical admissions tests. Judy has a special interest in writing assessment and research, and has managed the assessment of writing in a number of these tests including the creation of writing tasks, scoring rubrics, and marker training. She is currently part of a group working on a learning progression in writing.

Abstract

This presentation outlines different models of writing assessment that have a proven track record in various programs delivered by ACER. These models include holistic scoring, and partial analytical scoring on a number of criteria. The writing prompts for these models assume extended pieces of writing. The presentation will also introduce a new and alternative model of writing assessment that ACER has implemented in several large-scale assessments. This model is built on shorter pieces of writing from each student, designed to give more accurate and diverse insights into students' proficiency in writing in a range of text types, within the constraints of limited assessment time. Each short piece is coded using a different marking scheme with customised criteria. The three models – holistic scoring, partial analytical scoring and customised criterion scoring – are designed for different assessment contexts, from high-stakes selection to formative use by classroom teachers.

Introduction to ACER models of writing

ACER designs and delivers many writing assessments, from high-stakes assessments for selective schools and medical courses, to diagnostic assessments for classroom teachers working with their students. ACER currently uses three main forms of marking for this diverse range of tests: holistic marking, partial analytic scoring, and customised criterion scoring.

With holistic scoring, a student response is rated as a complete unit against a prepared scale. A global impression score is awarded denoting the level of achievement. Analytic scoring involves evaluating a student response by breaking it down into constituent elements or attributes, and assigning a mark for each criterion. Scores are aggregated to provide a single scale score. Two broad types of analytical scoring used by ACER are described here: partial analytical scoring, and customised criterion scoring.

Holistic marking

As noted, holistic marking at ACER involves assigning a global impression score to a piece of writing, usually on a scale from 1 to 10.

The ACER holistic model of writing from which a global impression is derived has three main assessment foci:

- Thought and Content (the quality of the thoughts, feelings and ideas presented)
- Structure and Organisation (the shape, form and cohesion of the piece)
- Expression and the technical aspects of language (the quality of the language used to present the ideas).

The three aspects in the model are treated somewhat hierarchically. Thought and content are prioritised over and above structure and the technical aspects of language in this model. This is because we posit that writing is about the process of generating ideas: creating ideas, not just recalling previously learned information. Writing can demonstrate original and divergent thinking, and the generation of new and varied responses to a set topic is valued. This process values writing as a type of *thinking*, through the written generation of ideas. This type of assessment of writing gives primacy to writing as a type of *thinking*, with the text's technical qualities of secondary importance. One advantage of this approach is that test takers with a language background other than English (LBOTE) are not penalised because of their lack of familiarity with the English language. Markers are always encouraged to 'look through' any obvious LBOTE errors in language, to the thought and ideas contained in the responses.

In the ACER tests of writing that use this scoring method, students may write in whatever genre or style they judge will best allow them to show their writing ability. Length is not a criterion. This form of scoring is used in the main for selective and admissions tests, for example the Cooperative Scholarship Testing Program (CSTP), the Graduate Medical Admissions Test (GAMSAT) and the Special Tertiary Admissions Test (STAT), where it is important to discriminate among applicants towards the top of the scale. Holistic scoring in these circumstances incorporates double-marking and discrepancy procedures, to ensure fair and accurate marking.

Partial analytic marking

ACER uses partial analytic marking schemas in some projects, both at school and tertiary level. It has been used in the International Schools' Assessment (ISA) test of writing since the test's inception. This is an assessment of writing, mathematical literacy and reading administered in over 335 international schools. It is undertaken by students from Grades 3 to 10 (pre-COVID, about 100,000 students in each annual cycle). It is underpinned by the principles of cultural eclecticism and administered by teachers in international schools in English, using detailed administration guides, but marked centrally by ACER markers. The participating schools receive reports at the school, class and student level and the students' results are used formatively to inform teaching and learning. Schools also use the assessment data to track progress over time and across year levels.

The writing component of the ISA comprises two extended writing tasks: one narrative/reflective task and one exposition/argument task. All students from Grades 3 to 10 write on the same two topics. The ISA Writing Task is based on historical development at ACER over several decades, of the concept of writing and the way students develop as writers. The marking and reporting scheme evolved for the ISA is very similar in nature to that used in the International Baccalaureate Middle Years Programme,¹ and the AERO (American Education Reaches Out) standards for writing.² The criteria for scoring also reference national examples of writing frameworks such as the McRel Standards,³ the Alberta writing program,⁴ the NWEA Six Traits,⁵ and the Australian Profiles for English and the New Zealand Curriculum.

For the ISA Narrative/Reflective task, students are asked to write a story or a reflective piece. The stimulus is usually a picture. The same prompt is used for all grades.

The ISA Exposition/Argument task requires a piece of writing that sets out ideas about a proposition. A few sentences or a short dialogue are provided as a prompt. Students may take an explanatory approach (exposition), a persuasive approach (argument), or they may combine the two approaches.

The ISA writing task uses a partial analytic scoring rubric. The criteria for assessing the Narrative/Reflective Task are:

- *Content* – broadly, the quality and range of ideas presented
- *Language* – sentence and paragraph structure, vocabulary and punctuation
- *Spelling* – phonetic and visual spelling patterns and the kind of words attempted, as well as correctness of spelling.

The criteria for assessing the Exposition/Argument Task are:

- *Content* – the depth and range of ideas presented, and the quality of reasoning
- *EAL Language* – the grammatical correctness and command of English syntax, sentence fluency and variation, and vocabulary
- *Structure and Organisation* – global and local organisation, and internal coherence.

Each student's writing thus receives six separate scores (that is, for three criteria on each of two tasks), which are then aggregated to give two scaled scores for writing. This approach to assessing writing is that it gives two kinds of reporting: a summative assessment of an individual's writing ability, which is mapped over time (from Years 3 to 10), and can be compared to others in the same year level both within schools and across schools; and important formative information for teachers to help them focus on their own teaching strategies.

Analytic scoring is also used in eWrite, an ACER online writing assessment which incorporates 10 different criteria across narrative, descriptive, informative and persuasive genres. In this assessment, because it is marked automatically using 'trained' algorithms, there is an emphasis on the technical aspects of student writing.

1 <https://www.ibo.org/programmes/middle-years-programme/curriculum/language-and-literature/>

2 http://projectaero.org/aero_standards/ELA/AERO-ELA-Framework.pdf

3 <https://www2.mcrel.org/compendium/topicsDetail.asp?topicsID=1013&subjectID=7>

4 <https://education.alberta.ca/media/160360/ela-pos-k-9.pdf>

5 <https://educationnorthwest.org/traits/traits-rubrics>

Customised criterion scoring

A third type of writing assessment design that ACER has pioneered in recent years is customised criterion scoring. The idea for developing this new form of writing assessment emerged from the recognised need to accurately assess what a student knows, understands and can do as a writer across a range of contexts. In more traditional forms of large-scale writing assessment, a very limited sample of a student's writing proficiency is collected: one extended piece of writing, or at most two, on which a judgement of proficiency is made. This means that the test taker has only one opportunity to demonstrate proficiency. Even when a partial analytical scoring model with several criteria is used – in some programs, such as NAPLAN writing, the single artefact is marked up to 10 times on specified criteria – the inescapable problem is that the same artefact is assessed repeatedly. The psychometric dependency among criteria is well-evidenced and probably unavoidable. We would not think of judging a student's maths or reading proficiency based on one item, would we? A further downstream problem of the single prompt instrument in a high-profile assessment is that it can have a stultifying effect on writing pedagogy, especially when a very limited and predictable range of text types is administered.

The key differentiating feature of the customised criterion scoring writing design is that each student is administered several writing tasks, within a manageable testing window. The assessment gives students the opportunity to demonstrate their proficiency across several different writing tasks, drawing on a range of writing skills both ideational and technical, covering different text types. Each task is marked on a different set of criteria drawn from an agreed framework but customised for the individual task. This is analogous to the way a reading or mathematics assessment is developed: each task reflects specific aspects of the framework.

The assessment framework for this kind of writing assessment comprises three dimensions or perspectives: content (text type – narrative, expository, instructional); context (the situation in which the writing is relevant, for example personal or 'wider world'); and process (e.g. generating ideas, structuring at micro and macro levels, syntax). Each writing task reflects one variable from each of the three dimensions. An important feature of the test design, or blueprint, is ensuring that, with regard to the process dimensions, there is a balance between the conceptual and the technical.

An individual student might be administered a set of three short tasks such as

- describing a picture in a couple of sentences (text type: descriptive)
- writing a quick note to a friend to arrange a play date, using details provided (text type: transactional)
- creating a half-page story based on a short textual prompt (text type: narrative).

Depending on the age and proficiency level of the students, at one end of the scale, a task might involve simply labelling a picture of an object (text type: label); at the other end, a task might ask the test taker to generate an argument of half to one page on a global issue (text type: persuasive).

Anything between two and five criteria are applied to a task, with the selection, number of scoring categories and specifics of the category descriptions customised to reflect the demands of the individual task. For example, a 'generate ideas' criterion with several scoring categories could be applied to each of the narrative and argument tasks sketched above, but would not be appropriate to the transactional or labelling tasks.

The multi-task writing assessment design, with customised criterion scoring, has been implemented by ACER in the New Zealand Literacy and Numeracy for Adults Assessment Tool (2010 to present), the Monitoring Trends in Educational Growth program (Afghanistan 2012) and in the regional assessment program, Southeast Asia Primary Learning Metrics (SEA-PLM 2019). This design is especially appropriate for sample assessment programs, in which each sampled student is administered a sample of the writing tasks, but it could also lend itself well to an adaptive writing instrument, using a combination of automated and human scoring.

Horses for courses

The three forms of writing assessment described all have their place. A single, holistic score captures an on-balance judgement about a piece of writing and has been shown, in analysis of programs including the writing assessments of the Western Australian Monitoring Standards in Education, and the Australian Cooperative Entry Program to be a reliable and efficient method of crystallising writing proficiency in a single score. It is appropriate where that is the main object, in selective entry programs such as those for scholarship selection or medical entry. Partial analytical scoring is suitable for combining the purposes of providing some formative analysis of students' proficiency in different aspects of the writer's craft, for teachers' use, with aggregated scaling of results for school or system consumption. Customised criterion scoring is particularly suitable for sample monitoring programs, and for reflecting a developed writing curriculum, where a wide range of text types and writing processes can be incorporated to give a comprehensive view of writing proficiency. Potentially, this kind of writing instrument is applicable to adaptive writing assessments, tailoring the type and difficulty of tasks to the proficiency of the individual test taker, and using a combination of automated scoring and human judgement.

In summary, each form of writing assessment and scoring has its virtues and deficiencies. A consideration of the purpose and context of the assessment will determine which type of writing instrument is most fit-for-purpose.

Reimagining classroom assessment and feedback to meet learner needs

Dr Fabienne Van der Kleij

Australian Council for Educational Research

<https://doi.org/10.37517/978-1-74286-685-7-5>

Dr Fabienne Van der Kleij is a Research Fellow at ACER. She leads ACER's research work in school and system improvement at the Centre for School and System Improvement. Her presentation will be based on her work as Research Fellow at The University of Queensland and Australian Catholic University, where she conducted research into classroom assessment and teacher and school leader professional learning in feedback to improve student learning. Fabienne's main research interests are educational assessment, school improvement, and feedback. Her most recent research has focused on teacher and student perceptions of classroom assessment feedback, student engagement with feedback, and assessment for learning in classroom practice.

Abstract

The power of classroom assessment and feedback to improve student learning outcomes has long been recognised. Yet, decades of research have yielded disappointing and often conflicting outcomes. This presentation challenges traditional conceptions of classroom assessment and feedback as teacher-driven practices. To meet learner needs better, it proposes a student-centred perspective in which students are active and have agency. By drawing on an extensive study of feedback reviews, this presentation illustrates how conceptualisations of feedback have changed over recent decades. This paper provides key insights into how classroom assessment and feedback practices can be changed in ways that are sustainable, afford student agency, and enhance student learning outcomes. Key learnings from recent research in primary and secondary education contexts are synthesised to provide state-of-the-art insights for shaping policy, practice and future research.

Introduction

The power of classroom assessment and feedback to improve student learning outcomes has long been recognised (e.g. Black & Wiliam, 2009; Hattie & Timperley, 2007). Whereas classroom assessment can pinpoint how students are going in relation to a particular goal, feedback is a critical vehicle to suggest next steps for improvement. As such, feedback is crucial in paving the way forward.

Despite its evidence-based potential, the effects of feedback on student learning in classroom assessment are often disappointing (e.g. Shute, 2008). Moreover, the effects of the same feedback on different students' learning outcomes are highly variable (Lipnevich et al., 2016). This suggests that individual students have different feedback needs, and assessment and feedback often do not meet these needs.

Classroom assessment is often constructed as something teachers undertake, with limited consideration of student roles (Hargreaves, 2012; Gamlem & Smith, 2013; Van der Kleij et al., 2019). Similarly, much of the feedback research has focused on identifying teacher-provided feedback that is generally effective for most students without considering how individual students respond to and act on feedback, and their agency in this process (Van der Kleij et al., 2017; 2019). Feedback can only enhance student learning if it is meaningfully embedded in classroom assessment practices and

students can use it in a productive way. Unfortunately, research shows that there are many potential barriers to effective feedback use. For example, students may not recognise feedback, may not understand it, may not perceive it as useful, or may not have the opportunity, ability or motivation to act on it (Gamlem & Smith, 2013; Van der Kleij, 2019; Van der Kleij & Lipnevich, 2021).

It is obvious that if we are to realise the power of classroom assessment and feedback, we need to reimagine classroom assessment and feedback in a way that meets learner needs. To begin to understand what this might look like, this article synthesises evidence-based findings from recent research in primary and secondary education contexts.

The purpose of feedback and role of the student

Feedback can take many different forms, and can be written, oral, or nonverbal communication. Research suggests that students do not necessarily recognise all such forms of feedback provided by teachers (e.g. Van der Kleij, 2019; 2021). Having conversations *about* what feedback is and its purpose can help students recognise and appreciate teacher feedback, and can clarify expectations and prevent miscommunications. When teachers make explicit their expectations for students to take an active role in feedback, the students are more likely to feel empowered (Van der Kleij et al., 2017).

Research suggests that students have different ideas about the purposes of feedback (Van der Kleij, 2021). For example, some students may regard feedback as corrective information and praise, whereas others identify feedback as taking place in any interaction with their teacher in which information is exchanged and common understandings are shaped. These differing perceptions of feedback are underpinned by students' perceived roles of themselves and the teacher in the feedback process. For example, some students see themselves as receivers of feedback, and teachers as providers (Van der Kleij, 2021). In contrast, students may see themselves as having an active role in assessing their own or their peers' progress, seeking clarification or further feedback, or provide feedback to themselves, peers and the teacher (Brooks et al., 2021b; Van der Kleij, 2021).

A recent study of feedback reviews published over the past 50 years (Van der Kleij et al., 2019) examined how the role of the student in feedback processes has been conceptualised. Key results are summarised in Table 1. The findings showed that although reviews have evolved towards a student-centred perspective (communication and dialogic models), the information processing model of feedback in which students have a limited role is still driving thinking in the field.

Table 1 Four models of feedback as distinguished by Van der Kleij et al. (2019).

Model	Description	Key features
transmission model	no student role	Students are passive and react to feedback in a predictable way.
information processing	limited student role	Students process feedback information in different ways based on their individual characteristics.
communication	some student role	Students choose how to make sense of feedback in interactions with others and self-generated insights.
dialogic	substantial student role	Students actively assess, seek, provide, receive and use feedback through self-regulated learning. Feedback effects are unpredictable and depend on various social, contextual and individual variables.

The most extreme way to reimagine classroom assessment and feedback is reflected in the dialogic model. In this model, teachers and students come to shared understandings of feedback by clarifying expectations, making one's own understanding explicit, and negotiating ways forward (Carless, 2013). What is key here is that teachers do not assume students understand feedback and know how to act on it, but follow-up to check on student receipt and understanding of feedback. In case of written feedback, this means that feedback is not seen as an endpoint, but rather, is used as an entry point for further discussion. In a classroom situation, dialogic feedback involves discussions with one or multiple students to gradually build shared understandings through multiple feedback loops. Importantly, the teacher and students all take on the roles of assessors and feedback seeker and provider (Van der Kleij, 2021; Van der Kleij & Adie, 2020; Van der Kleij et al., 2019).

The next sections in this paper focus on the key features of a student-centred perspective on classroom assessment and feedback. Research has demonstrated that if implemented with fidelity, these feedback strategies can substantially lift student achievement outcomes (Brooks et al., 2021a).

Clarifying what success looks like

A critical first step for the design of assessment is identification of learning outcomes or standards and what different levels of quality look like (e.g. Black & Wiliam, 2009). This step aligns the assessment and feedback process, as it sets the foundations by answering the question 'where am I going?' (Hattie & Timperley, 2007). Thus, if students are to take an active role in assessment and feedback, they need to understand what quality looks like. This can effectively be achieved through co-constructing criteria (Brooks et al., 2021a, 2021b).

Assessing progress

There are various formal and informal assessment methods that may be used to determine what progress students are making in relation to the learning goals. This step pinpoints how students are progressing, and aligns with the question 'how am I going?' (Hattie & Timperley, 2007). Teachers play a critical role in monitoring the progress of all students, but often run into issues of feasibility given large class sizes. Empowering students to assess their own or their peer's learning as a means to complement teacher assessment is therefore vital in driving sustainable assessment and feedback practice (Brooks et al., 2021a, 2021b). This requires an understanding of what success looks like, as well as an understanding of how to identify next steps.

Determining the best way forward

The true power of assessment and feedback lies in its potential to direct future learning. Once it has been established what the goals are and how learners are tracking, it is necessary to answer the question 'where to next' (Hattie & Timperley, 2007) to drive progress. Ensuring calibration of teacher or peer and self-assessment is vital here, as students are more likely to take action on feedback they have internalised and matches their self-assessment (Hattie & Timperley, 2007; Brooks et al., 2021b). This is particularly relevant given the potentially powerful (negative) emotional impact of feedback on students (Lipnevich et al., 2016).

Research shows that students prefer feedback that is specific (Van der Kleij & Lipnevich, 2021). However, feedback that is too specific may be counterproductive. If feedback is too directive in nature, thus telling students what to do, their autonomy in the feedback process is undermined. It only requires students to process the feedback at a relatively superficial level and does not require them to self-regulate or actively think about the feedback (Hargreaves, 2012; William 2011). Of course, it can be helpful to suggest ways forward, but leaving room for students' decisions in how to act on feedback empowers them in their learning, and makes it more likely that the feedback will be useful beyond the immediate task. However, highly specific or scaffolded feedback may be needed if the student is a novice learner, or if the student becomes frustrated because they cannot envisage how to proceed (Shute, 2008).

How can I reimagine my classroom assessment and feedback practice?

If this article has inspired you to apply these evidence-based findings to your classroom assessment and feedback practices, the following questions may provide a useful starting point for reflection.

- What do I believe to be the purpose of feedback and how do I see the student role in feedback?
- How can I help students develop more sophisticated understandings of what quality looks like?
- Which conditions need to be in place for peer feedback to be beneficial to students?
- What resources could I leverage to enable greater student autonomy in the assessment and feedback process?
- To what extent are my students capable of planning next steps for learning? How can I support them in this process, while empowering them as independent learners?
- How do I know that feedback has been impactful?

References

- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment Evaluation and Accountability*, 21(1), 5–31. <https://doi.org/10.1007/s11092-008-9068-5>
- Brooks, C., Burton, R., van der Kleij, F., Ablaza, C., Carroll, A., Hattie, J., & Neill, S. (2021a). Teachers activating learners: The effects of a student-centred feedback approach on writing achievement. *Teaching and Teacher Education*, 105, 103387. <https://doi.org/10.1016/j.tate.2021.103387>
- Brooks, C., Burton, R., Van der Kleij, F., Carroll, A., Olave, K., & Hattie, J. (2021b). From fixing the work to improving the learner: An initial evaluation of a professional learning intervention using a new student-centred feedback model. *Studies in Educational Evaluation*, 100943. <https://doi.org/10.1016/j.stueduc.2020.100943>
- Carless, D. (2013). Trust and its role in facilitating dialogic feedback. In: D. Boud & L. Molly (Eds.), *Feedback in higher and professional education: Understanding it and doing it well*. (pp. 90-103). Routledge. <https://doi.org/10.4324/9780203074336-11>
- Gamlem, S. M., & Smith, K. (2013). Student perceptions of classroom feedback. *Assessment in Education: Principles, Policy & Practice*, 20, 150–169. <https://doi.org/10.1080/0969594X.2012.749212>
- Hargreaves, E. (2012). Teachers' classroom feedback: Still trying to get it right. *Pedagogies: An International Journal*, 7, 1–15. <https://doi.org/10.1080/1554480X.2012.630454>

- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. <https://doi.org/10.3102/003465430298487>
- Lipnevich, A. A., Berg, D. A. G., & Smith, J. K. (2016). Toward a model of student response to feedback. In G. T. L. Brown & L. R. Harris (Eds.), *The handbook of human and social conditions in assessment* (pp. 169–185). Routledge. <https://doi.org/10.4324/9781315749136-17>
- Shute, V. J., (2008). Focus on formative feedback. *Review of Educational Research* 78(1), 153–189. <https://doi.org/10.3102/0034654307313795>
- Van der Kleij, F. M. (2019). Comparison of teacher and student perceptions of formative assessment feedback practices and association with individual student characteristics. *Teaching and Teacher Education*, 85, 175–189. <https://doi.org/10.1016/j.tate.2019.06.010>
- Van der Kleij, F. (2021). Teacher and student perceptions of oral classroom feedback practices: A video-stimulated recall study. *The Australian Educational Researcher*. <https://doi.org/10.1007/s13384-021-00478-0>
- Van der Kleij, F. M., & Adie, L. A. (2020). Towards effective feedback: an investigation of teachers' and students' perceptions of oral feedback in classroom practice. *Assessment in Education: Principles, Policy and Practice*, 27(3), 252–270. <https://doi.org/10.1080/0969594X.2020.1748871>
- Van der Kleij, F. M., Adie, L. E., & Cumming, J. J. (2017). Using video technology to enable student voice in assessment feedback. *British Journal of Educational Technology*, 48, 1092–1105. <https://doi.org/10.1111/bjet.12536>
- Van der Kleij, F. M., Adie, L. A., & Cumming, J. J. (2019). The student role in feedback: A meta-review of the literature. *International Journal of Educational Research*, 98, 303–323. <https://doi.org/10.1016/j.ijer.2019.09.005>
- Van der Kleij, F. M., & Lipnevich, A. A. (2021). Student perceptions of assessment feedback: A critical scoping review and call for research. *Educational Assessment, Evaluation, and Accountability*, 33, 345–373. <https://doi.org/10.1007/s11092-020-09331-x>
- Wiliam, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37(1), 3–14. <https://doi.org/10.1016/j.stueduc.2011.03.001>

Using PAT data to inform teaching and learning

Marc Kralj, Rachel Felgate, Dr Shani Sniedze-Gregory and Caithlin Power

Australian Council for Educational Research

Grant Barclay and Darren Leech

Marist College Canberra

<https://doi.org/10.37517/978-1-74286-685-7-9>

Marc Kralj is the Educational Consultant for ACER's School Assessments Services unit. Marc has worked across areas within the government education sector (DECD/Department for Education SA) for over 30 years. He has a deep understanding of the implementation of assessment strategies, moderation and reporting, curriculum content and learning design. Marc has extensive knowledge in using data and evidence to support teachers in informing their teaching and learning practices in and across school contexts. Marc has held several leadership roles that have focused on special needs programs, information and technology, and data analysis. His current role allows him to work with schools specifically in the areas of Progressive Achievement Testing and support leadership teams and staff to improve teaching practices and the learning outcomes of students in schools.

Rachel Felgate is a Senior Research Fellow in the Educational Monitoring and Research division at ACER. Rachel has worked internationally with over 20 years' experience in the United Kingdom, Aotearoa New Zealand and Australia. She has diverse experience in managing projects and undertaking quantitative research and report writing. Rachel has worked on the statistical and psychometric properties of leadership and teaching tools, led many evaluations of professional learning programs, led the Catholic Education South Australia (CESA) Professional learning in STEM, Learning Technologies and Data Analytics: Needs Analysis project and was the evaluation data and evidence lead on a seven-year professional learning and development contract in Aotearoa on behalf of a consortium, Te Toi Tupu, with delivery across English- and Māori-medium schools. Rachel has delivered many professional learning sessions for teachers and school leaders on using statistics, data and quality evidence of student achievement to support improvement.

Caithlin Power is the Project Director of System Clients within ACER's School Assessments Services Unit. Her role supports system clients (such as Departments of Education and Catholic Archdioceses) to implement ACER's Progressive Achievement approach across those networks of schools. Before she joined ACER, Caithlin worked as a primary school teacher and leader of eLearning for 11 years. She is experienced in administering Progressive Achievement Tests in the classroom, and using and interpreting data to understand students' learning abilities and to inform teaching and learning practices to improve student learning outcomes. Caithlin is committed to providing professional learning support to classroom teachers and school and system leaders and is passionate about engaging them in through meaningful, timely and important content.

Dr Shani Sniedze-Gregory is a Research Fellow in the Educational Monitoring and Research Division at ACER. She has 25 years' experience in education as a classroom teacher, curriculum developer and researcher, and has worked in both Australian and international contexts. Shani currently works on a range of research projects requiring qualitative data collection, as well as literature reviews, case studies, and program evaluation, and is on the ACER PISA 2025 team. Shani is particularly interested in the transfer of educational research into practical classroom resources and her projects often focus on working closely with practising teachers to enable these transformations.

Abstract

Using data to inform teaching and learning has become a common phrase in the education sector. What does 'using data to inform ...' mean and what does it look like in practice? This session delves into this question to illuminate multiple ways that data can be used to inform teaching and learning practice, and to demonstrate this through one school's journey so far. Our presentation is organised by three perspectives – Progress, Champions, and Research. 'Progress' addresses the nature of the suite of PAT assessments and resources, which build upon educational research for direct application in monitoring progress within the classroom context. 'Champions' showcases the way one school has organised its teaching and leadership structures to facilitate exemplary use of the PAT suite of assessments and resources. 'Research' provides a snapshot of how ACER focuses on the transformation of educational research into classroom practice, to make research more easily applicable in a classroom setting.

Progress

Educational data allow us to make informed decisions in the classroom. The processes of generating and accessing data have increased but the time and space needed to transform that data into a form that can be used in the classroom remains a key challenge for schools and teachers.

Educational data can be highly qualitative through to highly quantitative, and it can measure or reflect different domains, for example, subject content knowledge, synthesis and application skills, as well as timetable commitments and attendance. Some domains are better assessed qualitatively, some quantitatively, and some a mixture of both. ACER's Progressive Achievement Tests (PAT) are firmly in this mixed-methods category.

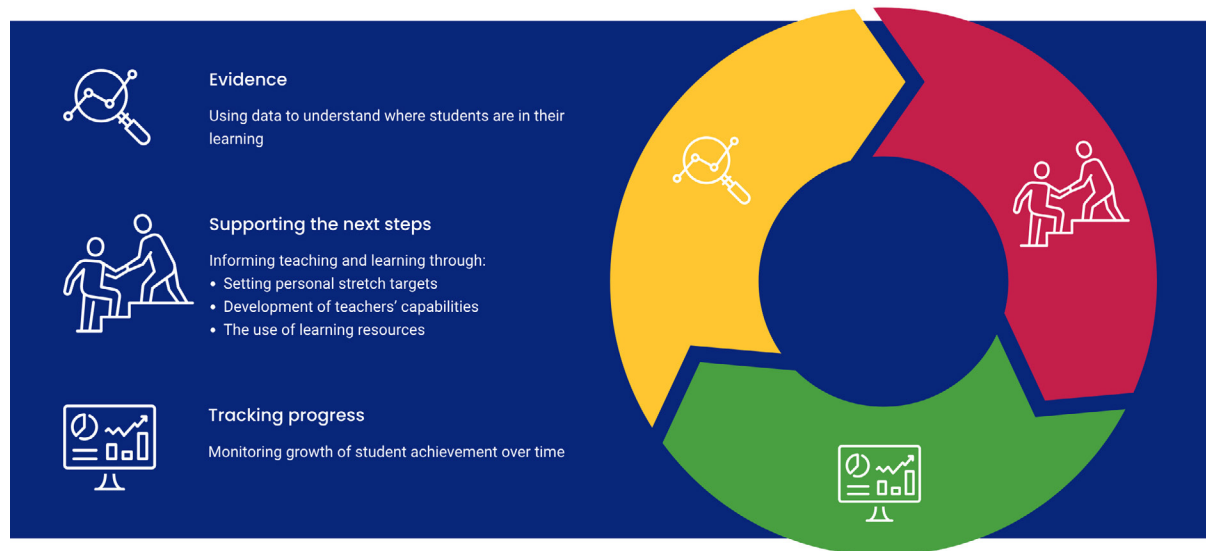
PAT assessments measure what students from Foundation to Year 10 know, understand and are capable of across subject domains, and help monitor progress over time. Innovative computer adaptive versions of PAT are also available: these assessments diagnose students' starting points to establish what students are capable of, then challenge them at the appropriate level for effective assessment. Through using PAT assessments over time, teachers can then monitor student progress and, when administered at recommended intervals, results can be used to measure learning growth. PAT assessments are flexible, secure and easy to administer; each assessment can be completed in approximately 45 minutes through a secure online platform. This also means that students can be assessed at a time appropriate to their needs, rather than being driven by external schedules.

ACER works with thousands of Australian schools to set up PAT functionality, assign and administer tests, and to assist teachers to understand and interrogate the resulting test data and act upon findings. To do this, ACER learning consultants discuss and plan all professional learning sessions collaboratively with each school they work with. Consultants ask schools to survey all staff to gauge a starting point for their learning. Like our students in a classroom, teachers also work at vastly different entry points, have varied experiences, and have quite distinct roles based upon learning domains and year-level experiences. Surveys allow for session planning that is relevant for teachers, no matter what their level of understanding or knowledge.

ACER's Progressive Achievement approach

Professional learning sessions focus on ACER's Progressive Achievement approach. This cyclical approach follows three stages, as shown in Figure 1.

Figure 1 ACER's Progressive Achievement approach (ACER, 2022)



The three stages of the Progressive Achievement approach are:

Stage 1: Collecting and using the evidence

Classroom work, quality assessment instruments such as progressive achievement tests, and our own expert knowledge combine to produce a wealth of data that tell stories about students' abilities. These data tell us what students know, understand, and can do at a particular point in time, and what they are ready to learn next.

Stage 2: Supporting next steps

The evidence and understanding that educators collect informs the next steps in teaching and learning by shaping personal stretch targets for every student. This process needs to be supported by learning resources and targeted support for a) students who are yet to master certain skills, b) students consolidating their current skills, and c) students requiring extension. At the same time, this approach helps develop and strengthen educators' own capabilities and confidence as professionals who effectively use data to identify and meet students' immediate needs.

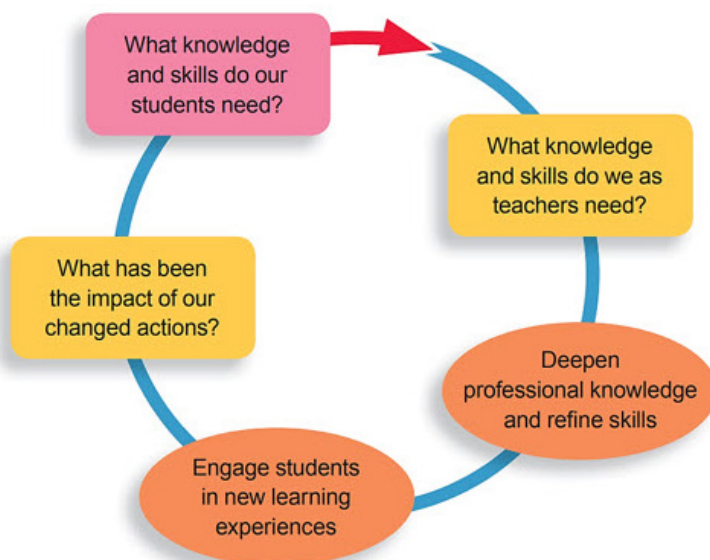
Stage 3: Tracking student progress

Monitoring student achievement over time is essential in order to see progress. The regular practice of collecting and reflecting upon data, including classroom work, observations, and results from reliable assessment tools continues to build the narrative of every student's learning and therefore track what progress has been made over select periods of time (recommended every 9–12 months for PAT assessments). This tracking also serves as a source of ongoing feedback to students and reporting to parents that supports further progress. A growth mindset approach in assessment, includes 'a belief that, regardless of where students are in their learning at any given time, every student is capable of making further progress' (Masters, 2016a).

Timperley's cycle of inquiry

ACER consultants also refer to Helen Timperley's 2011 cycle of inquiry and knowledge-building for teacher professional learning (see Figure 2). We incorporate this cycle of inquiry model to assist educators in their day-to-day practices. This model allows teachers to reflect on their teaching and learning and to use evidence when analysing student data.

Figure 2 Cycle of inquiry (as shown in Timperley, 2011)



The cycle begins by identifying, 'What knowledge and skills do our students need?' This information comes from curriculum documents, but also aligns with Stage 1 Collecting and using the evidence. PAT assessments provide information about what students already know within parts of the curriculum, and which parts might need further attention.

The second dimension of the cycle of inquiry supports and builds upon the first. Teachers are asked to reflect on their own knowledge and skills, their understanding of where their students are in their learning based upon the information collected, and understand that, 'in any given classroom, students are likely to be at very different points in their learning and development' (Masters, 2016b).

The third dimension phase, 'Deepen professional knowledge and refine skills', brings dimensions one and two together. Teachers can identify what will drive their own continued learning through professional learning opportunities – do they feel they have sufficient knowledge to act on the data gathered? Or, would they prefer to improve their skills in certain areas? In the context of PAT, this is where developing data literacy related to adaptive online assessments is critical, as well as how the data collected and interpreted relate to other formative and summative assessment practices, and how the data relate to what happens next in the classroom.

As a result of collecting evidence and engaging with professional learning opportunities, teachers are better supported to engage their students in new experiences; this is the fourth dimension of Timperley's cycle. These new learning experiences are tailored to build upon where students are in their learning and identify next steps in teaching and learning.

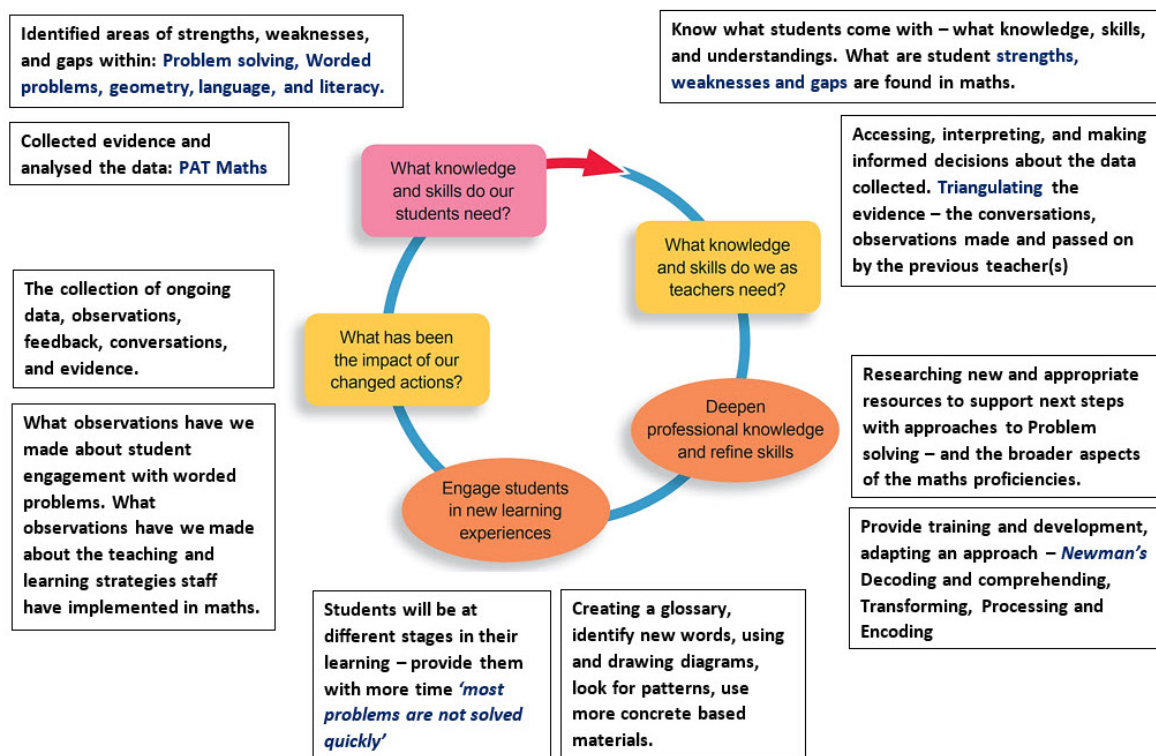
Changes in teachers' beliefs and knowledge through professional learning and development must result in some kind of change to teaching practices (Timperley, 2011, p.18).

The last dimension, 'What has been the impact of our changed actions?', encourages teachers to reflect upon the impact of the changes that have been made to teaching and learning resulting from earlier stages of the inquiry cycle. Surveys are a useful tool in allowing schools to access teacher feedback and to identify professional growth, just as they are for students. Surveys prompt professional reflection and discussion.

PAT linear and adaptive tests measure student knowledge and skills: this helps to identify learning needs and, when used at recommended intervals, demonstrate student progress over time. This can be interpreted as a direct reflection of the teaching and learning actions taken in between tests and can help in revising and replanning strategies for future teaching and learning.

Figure 3 shows how the Progressive Achievement approach (mathematics, in this example) can be aligned with Timperley's cycle of inquiry. It shows how a school might use evidence to support and direct next steps. For example, next steps in teacher professional learning, engaging students in new learning through targeted methodologies, including new or adapted resources, and gauging what impact change has had on the interventions.

Figure 3 Timperley's cycle of inquiry (2011), aligned with the PAT Mathematics approach



ACER consultants collaborate with schools to build confidence and capacity for all educators to:

- understand terminology and develop data literacies
- understand PAT scales and achievement bands to develop their knowledge of learning progressions within learning domains
- be 'hands-on' and involved in generating, interrogating, and interpreting the data they collect to make informed decisions in their teaching and learning – Achievement Reports, Item Performance Reports and Progress reports all provide the evidence that will inform teaching practices aimed at improving each student's learning outcomes
- be encouraged to work collaboratively in teams, faculty groups, professional learning communities (PLCs), and as a whole school.

Research

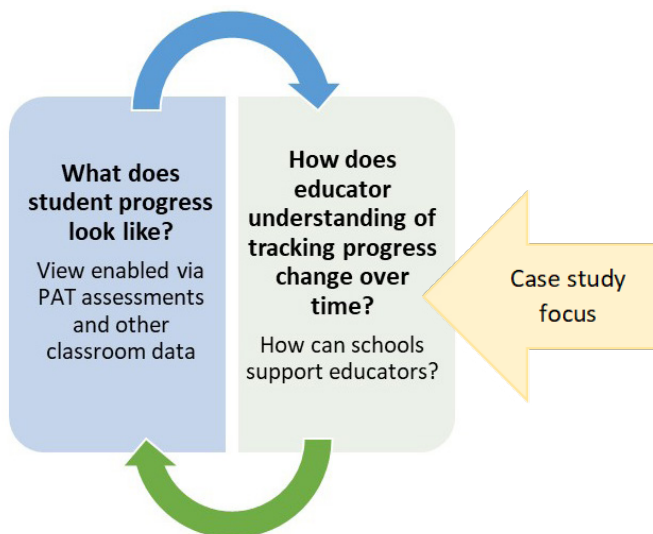
ACER is in the process of conducting case studies at selected schools who are using PAT products to improve their teaching and learning. We are recording these schools' PAT journeys as formalised case studies to share with other schools interested in how PAT assessments, data and resources can be used.

The case studies involve using secondary data from teachers through the surveys used by ACER PAT learning consultants. Primary data are also collected from a small number of teachers at these schools, through planned interviews or other feedback opportunities as they arise through the schools' normal use of ACER PAT products. Interview data collected relate to teachers' understanding of PAT practices, test data collection and use, and educational assessment in general.

We already know that through the use of the PAT suite of resources, student academic progress can be tracked over time. ACER researchers observe schools' practices in tracking student progress over time – a critical element in ACER's Progressive Achievement approach. The case studies also elevate these observations to the educator level (see Figure 4), namely, how educators' understandings of tracking student progress change over time, and what this means for using data in the classroom.

When completed, the case studies will be shared as a text (and potentially video) resource.

Figure 4 Case study focus



Champions

Marist College Canberra (MCC) is an independent Catholic school for 1700 boys in Years 4 to 12 and is one of ACER's case study schools. MCC has an embedded culture of academic rigour and high expectations and has the following goal for their students:

Our boys will be challenged intellectually every day and strive to make at least one year of growth every year (Marist College Canberra, 2021, p.8).

To achieve this goal the school has developed the Marist College Learning and Teaching Cycle (Figure 5), which represents their vision of teaching and learning. The cycle includes:

- future planning – 'Where to next?' – that identifies gaps and causes for those gaps
- deepening teachers' professional capacities to address gaps
- curriculum mapping and assessment design
- implementing a whole-school instructional framework
- using data to identify and measure growth and impact.

Figure 5 Marist College Teaching and Learning Cycle (Barclay & Stakelum, 2022)



MCC has also developed and implemented the Marist Learning Principles, which is a pedagogical framework that complements the Learning and Teaching Cycle (Figure 6). It supports MCC’s teaching and learning vision and reflects their teaching and learning approach. It incorporates six specific teaching and learning approaches:

1. cognition
2. targeted teaching
3. clarifying success
4. vocabulary instruction
5. responsive teaching and questioning
6. feedback

Figure 6 Marist Learning Principles (Barclay & Stakelum, 2022)



MCC’s overarching aim is for teachers to be able to use data to extend and lift the performance of every student in every class. Teachers use information from summative, formative, and external assessments to help identify the strengths and weaknesses of every student and then act upon this information. Challenging every student naturally requires significant support for teacher learning and development, including using grouping strategies and varied instructional material in the classroom.

Like many schools, MCC also understands the challenges of transforming assessment evidence into classroom teaching practice. The most pressing challenge they have found is the lack of time. Creating collaboration time for primary and secondary school levels, special teams, faculty groups, and leadership has led to the development of specific professional and development strategies. Redesigning staff meeting schedules to provide opportunities to discuss transforming data into teaching and learning strategies has been successful.¹

¹ These schedules will be discussed in the conference session.

More recently, Marist College Canberra has focused on next steps in supporting staff. Part of this has included the development of strategic teams, Champions, for maths and reading. These Champions are involved in training sessions to develop their understanding of PAT Adaptive Testing, with the aim of then sharing their knowledge and skills with the many teams across the primary and secondary schools. The Champion model is effective in allowing teachers to not only share their knowledge with colleagues, but to gain further experience in leadership practice.

Summary

The three elements of this research-to-practice endeavour, Progress, Research and Champions, coalesce at the school level and are enacted in the classrooms at Marist College Canberra. The combination of ACER progressive achievement practices, enabled in the MCC classrooms by teachers and supported by teacher leaders (Champions) are elaborated during the ACER Research Conference 2022. While this is one of the early opportunities for research dissemination, the ongoing case studies – with MCC and other schools around Australia – will exemplify data-supported teaching and learning practices and contribute to other schools' discussions on what might transfer to their own contexts.

References

- Australian Council for Educational Research [ACER]. (2022). *Progressive Achievement approach*. ACER. <https://www.acer.org/au/pat/progressive-achievement>
- Barclay, G. & Stakelum, L. (2022, May). Q&A: A school's perspective [Webinar]. ACER. <https://oars.acer.edu.au/help/webinars>
- Marist College Canberra. (2021). *Annual school report and school performance information 2021*. <https://www.maristc.act.edu.au/resources/publications>
- Masters, G. N. (2016a). *Growth mindset: Tracking student progress*. *Teacher Magazine*. Australian Council for Educational Research. <https://www.acer.org/in/discover/article/growth-mindset-tracking-student-progress>
- Masters, G. N. (2016b). *Monitoring student growth*. *Teacher Magazine*. Australian Council for Educational Research. https://www.teachermagazine.com/au_en/articles/monitoring-student-growth
- Timperley, H. (2011) *Using student assessment for professional learning: Focusing on student' outcomes to identify teacher's needs*. Paper No 21, May 2011. Department of Education and Early Childhood Development.

Sharing and securing learners' performance standards across schools

Emeritus Professor Richard Kimbell

Goldsmiths University of London, United Kingdom

<https://doi.org/10.37517/978-1-74286-685-7-6>

Richard Kimbell's specialist interest is in design learning and its interaction with assessment. He founded the Technology Education Research Unit at Goldsmiths University of London in 1990, and for over 25 years, ran research projects for research councils, for industry, for government departments, as well as for professional and charitable organisations. Richard has published widely in the field of technology education, including five single-authored books, has written and presented television programs and regularly lectures internationally. He has been a consultant to the National Academy of Engineering and the National Science Foundation in the United States, and a visiting professor at the University of British Columbia, the University of Stockholm, Edith Cowan University in Perth, Australia, Texas A&M University, and The University of the Shannon, Ireland.

Abstract

Assessing learners' performance makes very different demands upon teachers depending on the purpose and the context of the assessment. But common to all assessment is some sense of what 'quality' looks like. Most often teachers engage in formative assessments in the classroom, and the familiar standards of the classroom are adequate for this purpose. However if teachers are to undertake external, nationally regulated assessment then some sense of a national standard of quality is required. But there are very limited mechanisms by which teachers can acquire this understanding, so they use their best judgement, and standards vary from school to school not because anyone is attempting to cheat the system but simply because they cannot know what the real national standard is. It is for this reason that regulated examination bodies follow some process such as the following from the State Examinations Commission (SEC) in Ireland. '... teacher estimated marks will be subjected to an in-school alignment process and later a national standardisation process'. (SEC, 2021). How much simpler it would all be if teachers had – as a matter of normal practice – access to, and familiarity with, work from a national sample of schools, not just their own classroom.

Adaptive Comparative Judgement (ACJ) is an online assessment tool that has been used for some years, principally as a formative tool for learners (e.g. Bartholomew et al., 2018; 2019). This presentation reports on a study of the new ACJ Steady State tool from the same stable. The purpose of the new tool is to solve the problem of variable standards across schools by enabling teachers to make paired judgements of work from multiple schools and thereby evolve and agree standards of performance beyond their own school. The current study is operating in Ireland with a group of schools, a university, and the SEC. The anticipated outcomes include 1) better consistency of performance standards across schools in the research group and 2) greater understanding of and confidence in assessment judgements by the teachers. If ACJ has proved to be a powerful formative assessment tool for learners, ACJ Steady State is designed to be a formative assessment tool for teachers, helping to inform and support their assessment judgements.

Introduction

Teachers are constantly engaged in assessment activities, principally to better understand their learners' current performance. Mostly these assessment activities are informal and based around the on-going classroom tasks being undertaken by learners. Occasionally they have a more formal status, for example, when a teacher is asked to predict (by the school or perhaps by a parent) how well the learner is likely to perform in an examination. And even more occasionally, the teacher is asked to make a clear formal assessment of the standard of performance achieved by a learner as part of an external examination. As the teacher's judgements progressively migrate from informal-classroom to external-examination, they involve an ever-closer association with standards of performance outside the school; with those of other schools in the town; or in the next county; or even nationally. The problem for the teacher is that there is almost no mechanism by which they can know what those standards are. Examination bodies publish lists of assessment criteria and sometimes these are associated with exemplar materials, but every possibility cannot be covered and – in any event – quality cannot be defined in words (Polanyi, 1958). So inevitably the standards remain mysterious and open to interpretation and even misunderstanding.

While performance standards *within* a school can be carefully monitored and adopted by a teacher, it is almost impossible for that teacher to do the same thing *across* schools. This difficulty inevitably results in assessment error (The Office of Qualifications and Examinations Regulation [Ofqual], 2021); in teachers awarding different 'scores' for the same quality of work simply because of the lack of familiarity with the standards of work that apply in this school or that school and by the differences in the standards 'held' by the teachers as yardsticks of performance.

In simple terms, error is the difference between the result that a student ought to have been awarded from an assessment – given their level of attainment in the subject being assessed – and the result that they ended up being awarded (Ofqual, May 2021).

Comparative judgement for assessment

Between 2005 and 2010 at Goldsmiths, we ran project e-scape, a project commissioned by the Qualifications and Curriculum Authority (QCA) in England. The purpose was to explore an approach to performance assessment based on portfolios that could be created digitally by learners in the classroom and assessed digitally (online) by teachers and others. As part of that project, we created a new assessment tool based on the idea of comparative judgement (see Kimbell et al., 2009; Williams & Kimbell, 2012). The new tool, ACJ, was based on work by Laming (2004) who argued that all human judgements are comparative, and by Pollitt (2004) who first used comparative judgement in reliability studies for the University of Cambridge Local Examination Syndicate, a forerunner of Cambridge Assessment. The ACJ tool was used in prototype form in a pilot study in 2008 and as a developed tool in a national study across England in 2009. The results were encouraging, as Pollitt reported in the final e-scape report in 2009.

The final scale spread the portfolios out with a standard deviation of almost 3 units. The average measurement uncertainty for a portfolio was about 0.67 units, and the ratio of these two figures was 4.45. This means that the standard unit of the scale was almost 4.5 times as large as the uncertainty of measurement. This means the portfolios were measured with an uncertainty that is very small compared to the scale as a whole; this ratio is then converted into the traditional reliability statistic – a version of Cronbach's alpha or the KR 20 coefficient. The value obtained was 0.95, which is very high in GCSE terms (Kimbell et al., 2009, pp.73–81).

Quite apart from the reliability issue, however, what struck us most forcibly during the assessment process were the responses of the teachers undertaking the assessments. There were several hundred portfolios, all available online, and the ACJ system presented two at a time, inviting the teacher/judge to decide which was the stronger of the two. While the judge training involved the identification of headline criteria, these were not to be separated and scored but rather to be 'held-in-mind' as the judge came to a single holistic decision. Which is stronger ... this one or that one? All the teachers had supervised the e-scape classroom activity in their various schools and were familiar with the work. But – for the first time – they were able to see learners' work not just from their own school but also from schools all over the country. They were fascinated to see how different approaches to the work had developed and how their own students' performances related to others.

The ACJ tool has now become commonplace in schools, principally as a formative assessment tool. Asking learners themselves to make the paired judgements has the effect of promoting discussion of what 'quality' means in a piece of work, and teachers are adept at using such discussions to help enhance learners' performance. See for example the extensive collection of work by Bartholomew and colleagues in Utah in the United States (Bartholomew et al., 2018; 2019).

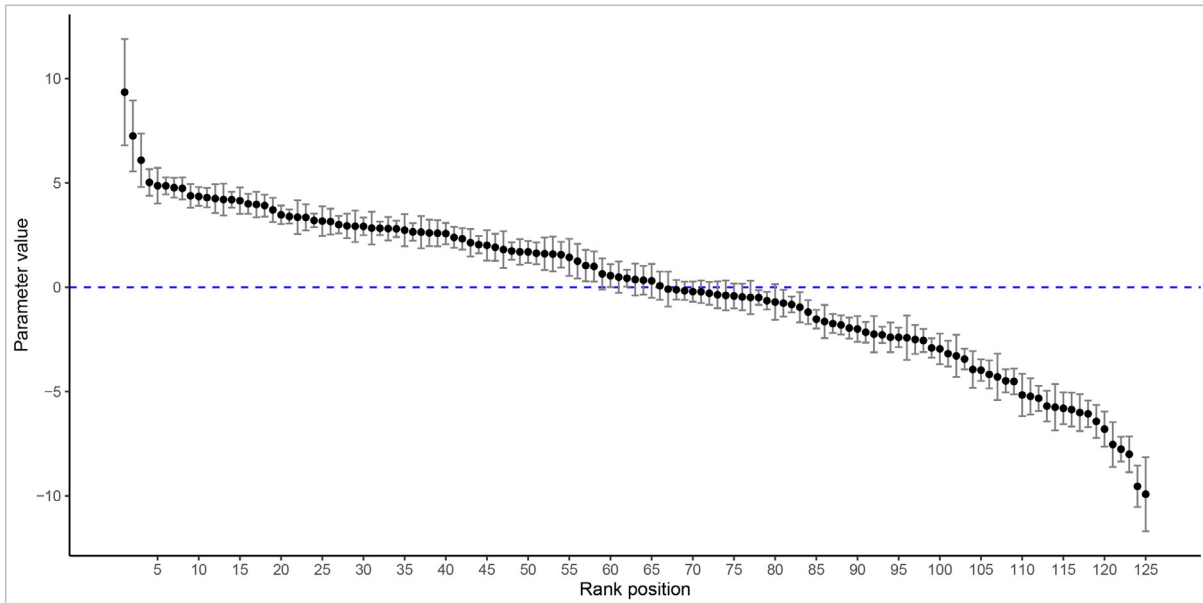
But, up to now, there has been a limitation with ACJ in that it is a single-cohort tool. So a class in School A can do an ACJ exercise, and another class in School B can do a different one. This will result in two performance ranks, but with no means to relate them. Of course the two schools could collaborate on a single bigger ACJ session, but that is much more complex to arrange and to manage for a class teacher. We came to see that it would be very useful if separate ACJ ranks could – in some way – be combined into a single rank. Thus was born the idea of a new form of ACJ that we have provisionally called 'ACJ Steady State'. This is being developed by RM Education (a leading supplier of learning and assessment resources to the education sector) who acquired and refined the original ACJ algorithm. (see <https://www.rm.com/>). In collaboration with RM, a pilot study is underway in Ireland involving the SEC and the Technological University of the Shannon. The study involves a trial with teachers and learners in 10 schools pursuing a graphics program at Leaving Certificate level.

A pilot study of ACJ SteadyState

If we imagine that a class of learners in four schools A,B,C,and D has each produced (say) 25 pieces of work, we can collect a small team of teachers to undertake a standard ACJ process on the 100 portfolios. This will result in a simple rank order of the 100 pieces.

For the purposes of this study, this rank then becomes 'the ruler', against which we might choose to measure other work.

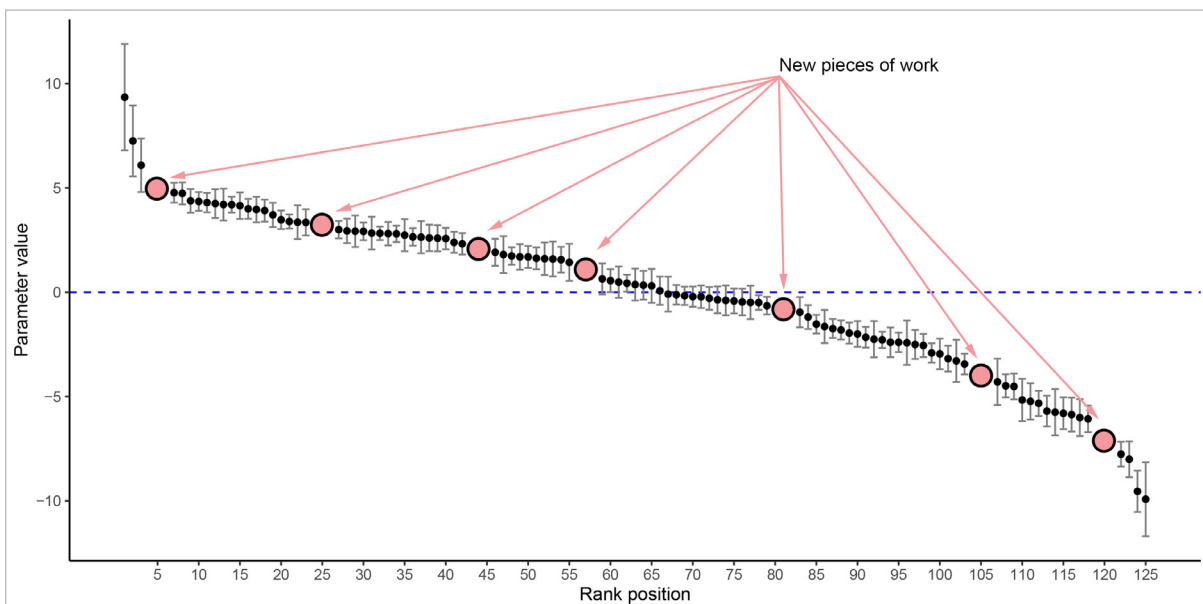
Figure 1 ACJ rank becomes 'the ruler'



If we also now imagine a fifth school (E) that has also undertaken the same exercise as the classes in schools A–D. This new work (another 25 pieces) will be added to the pot of portfolios and will be judged by the same team of teachers, but this time using ACJ Steady State.

It is important to recognise that in normal ACJ exercises, the positions of *both* of the portfolios will change as the result of a judgement. The rank continuously modifies itself as the judging unfolds. But ACJ Steady State operates differently. The ruler is fixed and the judging thereafter only adjusts the position of the new pieces until they reach their points of stability within the fixed rank.

Figure 2 ACJ Steady State with new work located in the ruler



Provided this all works as we believe that it will, then it will be possible to create a ruler with a selected number of schools, and subsequently we will be able to integrate new schools into the ruler. All it requires is that the judges make their judgements of the new school's work against the work in the ruler. ACJ Steady State is, in effect, a multi-cohort tool and we can go on and on adding schools. It will thus be possible to create not just school ranks but also regional ranks and even a national rank. It should be noted that the judging process for ACJ Steady State will be lean and quick. The algorithm will only select pairs involving a new portfolio, sometimes comparing it to a 'ruler' portfolio and sometimes to another new one. Moreover the new algorithm only has to calculate and adjust the position of that new piece. While the ruler itself (within ACJ) required approx 15 rounds of judging to fix the positions of the 100 pieces of work, our modeling suggests that (within ACJ Steady State) five or six judgements will fix the positions of the new work.

Issues to be explored

The issues to be tackled here include some that might be thought of as informal, formative issues for teachers and schools, while others concern more technical matters that have to be dealt with for formal assessments.

Formative classroom issues

1. It is critical to the purpose of the project that the teachers are empowered to discuss the judgements they make. Wherever possible teachers will work as pairs, agreeing the judgements as they go and noting the qualities of performance that led to the judgements. We believe that it will be possible for a group of teachers not just to operate reliably but also to articulate and agree as a group what the *qualities of work* are at all levels through the rank.
2. We want to explore the pedagogical implications of the ruler. When teachers are self-consciously aware of the standards of performance that make up the scale, what do they choose to do with that information? Might they choose to use ACJ Steady State as a regular progress check for themselves? Might they engage learners in using it to observe work beyond their own school? Might they choose to create rulers not just for overall performance but also for *elements* of performance?

Technical assessment issues

1. When schools (or the SEC) wish to create a ruler that reflects a 'national standard' it is important to consider how this might be done. Should this be from a random group of schools? How big should the ruler be? Should it include selected work from a previous year's performance? What happens if work from a new school goes off the scale at one of the ends? The project will enable us to explore and recommend a model of practice for deriving and applying the ruler.
2. Once a new school has been integrated into the ruler, it will be possible not only to comment on the placing of individual portfolios, but also to derive a 'school statistic' that identifies how, for example, the mean or median performance of School A is different from the mean or median performance in the ruler. This might enable a moderation tool to be developed to assist the SEC in their desired adjustment of teacher assessments 'teacher estimated marks will be subjected to an in-school alignment process and later a national standardisation process' (SEC, 2021).

Conclusion

Currently it is only at external examination times that teachers get the opportunity to see their learners' work set against a backdrop other than their own school's, and at that moment it is already too late for the information to feed into positive learning experiences. The existing ACJ tool is currently in use in classrooms where teachers use it as a formative feedback tool to help learners to get a better understanding of their own work. Our vision for ACJ Steady State is that it should be a *formative feedback tool for teachers*; enabling them to share work across multiple classrooms and debate standards across schools. We see this as a tool that will help to de-mystify national standards, helping schools to collaborate and allowing teachers to gain confidence through their shared judgements.

References

- Bartholomew, S. R., Strimel, G. J., Garcia Bravo, E., Zhang, L., & Yoshikawa, E. (2018). *Formative feedback for improved student performance through adaptive comparative judgment* [Paper presentation]. 125th ASEE Conference, Salt Lake City, Utah, United States.
- Bartholomew, S. R., Strimel, G. J., & Yoshikawa, E. (2019). Using adaptive comparative judgment for student formative feedback and learning during a middle school design project. *International Journal of Technology and Design Education*, 29(2), 363–385.
- Kimbell, R., Wheeler, T., Stables, K., Shepard, T., Davies, D., Martin, F., Pollitt, A., Whitehouse, G.. (2009). *E-scape portfolio assessment: a research and development project for the Department for Education & Skills (DfES) and the Qualifications and Curriculum Authority (QCA), phase 3 report*. Technology Education Research Unit, Goldsmiths, University of London.
- Laming, D. (2004). *Human judgement: The eye of the beholder*. Thomson.
- The Office of Qualifications and Examinations Regulation (Ofqual), (2021, 17 May). Bias in teacher assessment results. *The Ofqual Blog* <https://ofqual.blog.gov.uk/2021/05/17/bias-in-teacher-assessment-results/>
- Polanyi, M. (1958). *Personal knowledge: Towards a post-critical philosophy*. Routledge & Kegan Paul Limited.
- Pollitt, A. (2004, September). *Let's stop marking exams*. [Paper presentation]. IAEA Conference, Philadelphia, United States. <https://www.cambridgeassessment.org.uk/images/109719-let-s-stop-marking-exams.pdf>
- State Examinations Commission [Ireland] (2021) *Leaving certificate 2021 and accredited grades*. https://www.citizensinformation.ie/en/education/state_examinations/leaving_certificate_2020_calculated_grades.html
- Williams, P.J., & Kimbell, R. (Eds.). (2012). Special issue on e-scape. *International Journal of Technology and Design Education*, 22(2).

Assessing reading: How assessment can be used to target teaching and enhance understanding of reading comprehension

Dr Sandra Knowles

Australian Council for Educational Research

<https://doi.org/10.37517/978-1-74286-685-7-7>

Dr Sandra Knowles is a Senior Research Fellow and Team Leader of the Humanities team in the Assessment and Reporting division at ACER. She has worked in test development since 2009, specialising in literacy for school-aged children. Sandra's work has primarily involved the development of national and international assessments and teaching resources across the skill areas of reading, speaking and listening, writing, and 21st-century skills. Currently, her work focuses on the development and implementation of learning progressions in reading to inform formative assessment practices and improve teaching and learning in the classroom. Prior to working at ACER, Sandra was an ESL teacher in English for Academic Preparation and completed a PhD in English at the University of New South Wales.

Abstract

The skills demonstrated by a proficient reader are not easy to untangle. Current research acknowledges that reading comprehension is a highly complex area of ability, one that needs to be understood as the coordination of a number of integrated processes. Using example test questions and data, this presentation explores how assessment can help us make sense of reading comprehension in a way that curricula and commonly used teaching strategies cannot. Assessment is evidence that informs us about the skills involved in the reading process, how they relate to each other, and how they develop in complexity. When assessment is understood in this way, as much more than a tool to compare a student's ability with that of their peers, it can be used to identify what skills individual students are consolidating, and what specific steps will help their development.

Introduction

The skills demonstrated by a proficient reader are not easy to untangle. Current research recognises reading comprehension as a highly complex area of ability, one that needs to be understood as the coordination of a number of integrated processes that are both important and changeable (Conley & Wise, 2011, p. 93; Kendeou et al., 2016, p. 63; O'Reilly et al., 2014, p. 404). Curricula commonly reflect this complexity by covering a wide range of reading components, such as knowledge of text structures, making inferences and evaluating authorial devices. Teaching strategies and resources, such as dialogic reading, think-aloud and reading responses, encourage students to think more deeply about what they are reading and may help teachers to broadly identify where students need support (e.g. whether they are better able to recognise literal than inferential meaning). This paper explores what assessment can offer to our understanding of students' reading ability and how it can support the use of curricula and resources.

Well-designed assessments provide valuable information

Standardised assessments allow teachers to identify where students are performing in a particular learning domain, but the question then becomes – performing in what? In unpacking the ‘what’, we learn about the integrated reading processes and help to make sense of them. While no assessment could capture the process of understanding what we read in its entirety, a comprehensive, well-designed assessment can provide valuable information about the specific skills and processes that constitute reading comprehension. This specificity serves to elaborate the broader descriptions provided in curricula and provides better targeting for teaching practices. It is not so much a matter of untangling reading processes, but of focusing on one at a time. Due to the integrated and holistic nature of reading comprehension, improvement in one area will inevitably influence others. Note that the term ‘skills’ is used here to refer to the broader abilities being demonstrated (e.g. locating, interpreting or evaluating) and ‘processes’ refers to the ways in which these broader abilities are being applied (e.g. making connections to interpret information).

Standardised assessments, like the Progressive Achievement Tests (PAT) and NAPLAN, are still commonly perceived as tools that rate a student’s ability in comparison with their peers. This is certainly something standardised assessments do, but as Dylan Wiliam explains, ‘any assessment is formative to the extent that evidence about student achievement is elicited, interpreted, and used by teachers, learners, or their peers to make decisions about the next steps in instruction’ (2011, p. 43). Higher student growth in reading comprehension is achieved when teachers have information about their students’ reading progress, and assessment that informs teaching is increasingly acknowledged as an effective way to ensure that instruction is timely and appropriate to students’ needs (Conley & Wise, 2011, p. 97; Förster et al., 2018, p. 99). Data-driven instructional decisions are now being encouraged across many schools and systems (Griffin, 2012, p. 2). In the report known as Gonski 2.0. Finding 7 refers to the compelling evidence that ‘tailored teaching based on ongoing formative assessment and feedback is the key to enabling students to progress to higher levels of achievement’ (Department of Education and Training, 2018). In monitoring students’ progress, teachers can identify where there is a need for extra support, adapt instruction as appropriate, and evaluate their approach to teaching the key components more generally. (Anderson & Primary English Teaching Association Australia [PETAA], 2016, p. 108; Förster et al., 2018, p. 98).

Diagnostic insights from assessment data

For teachers to make informed decisions about their students, they require assessment data that capture the skills and processes involved in reading comprehension, and in a way that they can relate to their practice. Assessment data are only diagnostic if they enable educators to identify where support is most needed. This paper explores the diagnostic potential of test questions, and specifically the insight they provide into where students’ understanding may be lacking, and the next steps for improvement. Take the following example of a test question.

Seru, Claire and Mike were playing at the edge of the river. Mike dipped his toes into the water, but Seru didn’t want to. He said it was too cold. Claire crept up behind Seru and gave him a little push.

What was Claire trying to make Seru do?

To answer this question, a student needs to connect information across the four sentences to infer that Claire’s intention was to push Seru into the river. While relating the first and last sentences may be enough to make the inference, recognising Seru’s reluctance to enter the water supports a more comprehensive understanding. Forming this understanding requires linking a pronoun reference across two of these sentences, which is an added complexity. If a student has not understood that

Claire intention was to make Seru fall into the river, they have not successfully made the link across sentences. If a teacher wanted to obtain further information about what this student *could* do, a logical step would be to see if they could make a simpler connection within or across a sentence. Take this second example.

Ian threw the ball to Meg. Crash! The vase lay in pieces on the floor.

What happened when Ian threw the ball to Meg?

This is an easier connection to make than in the first example. The sentences are shorter, so a student only needs to maintain the link between ideas across a small section of text, and there are no pronoun references. But a link still needs to be made, relating the action of throwing the ball to the broken vase. If the same student as before answers this question correctly, the teacher would know that they *are* able to make simple links but need to be supported to make more challenging links across longer sections of text, with added complexities such as pronoun referencing.

These examples provide some indication of the insight that can be gained from interrogating the skills and processes required to comprehend text. To answer these questions, a student is required to interpret information (skill) by making connections across adjacent sentences (process). These examples also illustrate that the process of making connections increases in difficulty. At the simplest level, it involves linking explicit ideas within a sentence or across simple sentences. For proficient readers, it could involve making links across an extensive piece of text, where ideas are subtle or technical or there is some other complicating factor. This progression is important; knowing where a student is at in their ability to perform this process is only useful if you have some idea of where they need to go next. As current research suggests, assessment in reading comprehension should support teachers in understanding the key components that underpin growth (Anderson & PETAA, 2016, p. 108; Förster et al., 2018, p. 98; O'Reilly et al., 2014, p. 404).

Such a deepening of teachers' understanding can have significant implications for classroom practice, in particular by informing their evaluation of a student's ability in a variety of common activities. Retell of a text (either through think-aloud or other approaches) is one such activity that is often used to evaluate student's reading comprehension skills. But when retelling or recalling what they've read in their own words, students are using a range of processes that may be challenging to distinguish. For example:

- Can they identify the main idea and differentiate it from the details?
- Can they logically sequence events?
- Can they link cause and effect to make sense of events?
- Can they rephrase ideas using different vocabulary and sentence structure?

Asking a student to retell what they have read is a useful task, but it can be challenging for teachers to know what they are looking for and how to support a student who may struggle with the task. As test questions tend to focus on one of these processes at a time, assessment data have the potential to inform which of these areas a student may have most difficulty with. Take the following example.

Geckos range from about 8 to 15 cm in length, have a long snout and a long, thin tail. Their skin colour can be grey or pinkish and some have patterns. They have soft bodies and scales. They use the suction pads on their toes to stick to walls and other objects.

What does this paragraph tell us about geckos?

- what they eat
- where they live
- what they look like
- how long they live for

By recognising that this paragraph is about what geckos look like, a student has been able to categorise the explicit clues in the paragraph by relating them to a simple summary. This student could be challenged in a variety of ways, such as by performing a similar task with less prominent clues and perhaps with a different text type. A student who is unable to answer this question has not been able to relate the clues to the summary provided and could therefore be given a much simpler related task, such as organising familiar items into categories (e.g. food, animals, activities).

A different kind of question, such as sequencing events in a narrative, would provide information about a different process. If a student sequences events in a linear narrative, they may have demonstrated the ability to recognise a summary of the events, or synonymous matches for key words, depending on the question. If the temporal structure of the text is not linear, the student may have interpreted some challenging conjunctival or adverbial phrases to recognise this. If they could identify the main idea of a text from a list that includes less important details from the text, they have demonstrated an ability to make this distinction. While all these tasks relate to the general skill of summarising, they each target a specific reading process that offers guidance for the teacher about 'where to next'.

Assessments are springboards to further investigation

Understanding these processes and their progression means understanding how student ability can be evaluated and supported. It means that there are steps that can be taken – steps based on evidence – to assist students who are currently identified as failing to meet the expected level of performance represented in curricula. Low levels of reading comprehension ability need to be caught early. Evidence suggests that students who enter secondary school with a low reading capacity will find it difficult to engage with the material taught in school and are therefore likely to fall further behind. Further, this struggle is likely to continue beyond their education into employment (Griffin, 2012, p. 2; Kendeou et al., 2016, p. 63; Woodford, 2016, p. 43).

Using assessment to learn about and support students in reading comprehension is not without its challenges. Time-poor teachers are unlikely to be able to interrogate every question in an assessment for every student, and there is no guarantee that this interrogation would be fruitful in every instance. Assessments are best considered as springboards to further investigation, and the data can be explored intermittently, depending on the teaching focus at the time. The data can, and must, be part of a process of ongoing formative assessment and feedback, rather than a conclusion about students' achievement levels. It can also be usefully organised to simplify the investigative process. Reading questions can be categorised not only according to the broader skill areas (e.g. locating, interpreting and evaluating), but also according to the dominant process required to answer the question (e.g. making connections, generalising and summarising, comparing and contrasting, recognising authorial devices and so on).

Conclusion

Assessment data can also assist teachers to identify the developmental sequence of typical skill acquisition by looking at the relative difficulty of items and the complexity of the texts. The examples provided in this paper illustrate the range and focus of reading skills and processes that underpin a quality described proficiency scale or learning progression. Assessment in reading comprehension should support teachers in understanding the key components that underpin growth (Anderson & PETAA, 2016, p. 108; Förster et al., 2018, p. 98; O'Reilly et al., 2014, p. 404). The Gonski report (Department of Education and Training, 2018) refers to the usefulness of evidence-based learning progressions in setting performance expectations, understanding and reporting progress, informing teaching and learning and promoting educator professional development. When the learning progression or described scale is the focus point, test questions and their texts serve as illustrations of the different levels. These descriptions would ideally encourage a shared language that extends current understanding of reading comprehension and elaborates curriculum statements (such as the references to 'comprehension strategies' across the Literacy strand at different year levels). (Australian Curriculum and Reporting Authority [ACARA], 2018). This deepening of understanding could be shared with students, who would come to better recognise their own development and have clear goals to work towards.

References

- Anderson, P. (2016). Assessment and growth in reading comprehension. In H. Fehring (Ed.), *Assessment into practice: understanding assessment practice to improve students' literacy learning* (pp. 108-114). Primary English Teaching Association Australia.
- Australian Curriculum and Reporting Authority (ACARA). (2018). *Australian Curriculum, English* (Version 8.4). <https://www.australiancurriculum.edu.au/about-the-australian-curriculum>
- Conley, M. W., & Wise, A. (2011). Comprehension for what? Preparing Students for Their Meaningful Future. *Theory Into Practice*, 50(2), 93–99. <https://doi.org/10.1080/00405841.2011.558411>.
- Department of Education and Training. (2018). *Through growth to achievement: Report of the review to achieve educational excellence in Australian schools*. Australian Government. <https://www.dese.gov.au/download/4175/through-growth-achievement-report-review-achieve-educational-excellence-australian-schools/18692/document/pdf/en>
- Förster, N., Kawohl, E., & Souvignier, E. (2018). Short- and long-term effects of assessment-based differentiated reading instruction in general education on reading fluency and reading comprehension. *Learning and Instruction*, 56, 98–109. <https://doi.org/10.1016/j.learninstruc.2018.04.009>
- Griffin, P. (2012, August 27). *The influence of teaching strategies on student achievement in higher order skills* [Paper presentation]. Research Conference 2012, Darling Harbour, New South Wales, Australia. https://research.acer.edu.au/research_conference/RC2012/27august/19
- Kendeou, P., McMaster, K. L., & Christ, T. J. (2016). Reading comprehension: Core components and processes. *Policy Insights from the Behavioral and Brain Sciences*. <https://doi.org/10.1177/2372732215624707>
- O'Reilly, T., Weeks, J., Sabatini, J., Halderman, L., & Steinberg, J. (2014). Designing reading comprehension assessments for reading interventions: How a theoretically motivated assessment can serve as an outcome measure. *Educational Psychology Review*, 26(3), 403–424. <https://doi.org/10.1007/s10648-014-9269-z>
- Woodford, H. (2016). Engaging your adolescents with reading. *Literacy Learning: The Middle Years*, 24(2), 43-54. <https://search.informit.org/doi/10.3316/aeipt.212348>

Teachers' assessment literacy and design competence framework

Dr Sladana Krstic

Australian Council for Educational Research, UK

Dr Sarah Richardson

Australian Council for Educational Research

Dr Sarah Manlove

The International Baccalaureate Organization

<https://doi.org/10.37517/978-1-74286-685-7-8>

Dr Sladana Krstic is a Senior Research Fellow at ACER UK where she leads the Research and Psychometrics team. Sladana has managed many projects, most recently for the Royal Society of Chemistry, the British Council, the International Baccalaureate, and many others. She is currently engaged in a groundbreaking research project for the International Baccalaureate Organization and the Jacobs Foundation on Curiosity, Creativity and Learning. She previously worked at the Office of Qualifications and Examinations Regulation (OFQUAL) in the United Kingdom where she undertook research on the development and implementation of the qualifications and skills programs. Sladana has over 15 years' experience in research and project management, and high-stakes assessment policy communications in the education sector. Sladana has a Doctorate in Education and is an expert researcher in quantitative and qualitative methodology, data analysis, sampling designs, and questionnaire development.

Dr Sarah Manlove is a Research Manager at the International Baccalaureate Organization who has specialised in policy, curriculum and program development research for IB's K-12 continuum for nine years. Her research background has focused on the design of student-centered supports for metacognition in technology-enhanced learning environments, and mechanisms for learning transfer. Sarah received her PhD in Instructional Design from the University of Twente, the Netherlands in 2007 and undertook post-doctoral work in cognitive science at Indiana University.

Dr Sarah Richardson is the Research Director for the Australian Council for Educational Research UK. In this role Sarah brings together her extensive experience in educational research (particularly in assessment [primary, secondary and tertiary], evaluation of educational projects, education policy and data collection) and desire to contribute to a better understanding of educational issues in the UK, Europe and beyond. Sarah has a PhD in education and politics from the University of Melbourne, a Master's Degree in International Relations (thesis on literacy programmes for women in South Asia) from the University of Amsterdam and a Bachelor's Degree in Human Geography from the University of Liverpool. With this background Sarah thrives in international projects and is fascinated by the intersection between education, culture, society and policy.

Introduction

Teachers play a central role in facilitating learning and empowering learners (Pastore & Andrade, 2019). They must possess a high level of assessment literacy to properly fulfil their role of implementing teaching and facilitating students' learning (Hattie, 2009). Without it, the quality of educational provision can be impacted (Looney et al., 2018).

The expectations placed on teachers in relation to their assessment knowledge and skills tend to focus on their ability to design and implement assessment with students, and to use assessment data to inform teaching practices (Brookhart, 2011). Much of the emphasis in the literature is on the mechanical processes involved in assessment, and on what is required by individual teachers (DeLuca & Bellara, 2013).

Our holistic approach conceptualises assessment literacy as relating to teachers' assessment competence, sound practices, socio-cultural understandings, and how these are embedded within the educational context (DeLuca et al., 2019). The focus is on the body of teachers within a school; many practicing teachers are at a novice stage in their professional practice (Xu, 2017) and teaching needs to be supported by school structures for optimal student learning (Crichton & McDaid, 2016).

Methodology

The aim of this study was to develop an assessment literacy and design competency framework for use across International Baccalaureate (IB) programs and to inform professional learning resources. The framework was designed to summarise the scope of assessment literacy that needs to be addressed within the IB.

The research undertaken in developing the framework comprised three elements:

- a literature review
- a scan of IB documents
- consultations with IB staff.

The framework provides a description, methodology (the attitudes, behaviours, knowledge, and skills related to each element of assessment literacy) and resources for each of the main elements identified. The resources are intended to be additional to IB documents and include exemplar courses, and other reference materials.

The framework is being used within the International Baccalaureate Organization to inform professional development for teachers and to inform curriculum and program design.

Key findings

Figure 1 shows the seven elements identified in the framework: assessment knowledge and skills; formative assessment; assessment identity; professional development; school environment; engagement of learners; and integrating digital assessment. Each element is essential individually, but also needs to work in conjunction with the other elements.

Figure 1 The seven elements of the assessment literacy and design competency framework



Assessment knowledge and skills

The knowledge, skills and attitudes incorporated in definitions of assessment literacy are influenced by changes in attitudes towards education and assessment over time (Yan & Cheng, 2015), and equally influenced by teacher career development (Coombs et al., 2018). Ideally, all teachers have sufficient skills and knowledge to evaluate the extent to which assessment materials can generate reliable and valid data (Choi et al., 2021). Across a school, the teaching body needs to be proficient in several fundamental areas of assessments:

- developing and grading rubrics for open-response tasks (Brookhart, 2011)
- using assessment data to monitor learner progress and to identify ways to enhance learning
- using assessment results in their teaching practices (Mertler, 2009).

Formative assessment

Formative assessment is regarded as the backbone of good teaching practice (Moss & Brookhart, 2009). It is important that all teachers are aware of the philosophies that underscore formative assessment and the wide range of approaches that comprise formative assessment, from questioning (Kyriacou, 2010) to observation, and to project work. In addition to providing opportunities for learners to demonstrate skills and knowledge in curricula areas, teachers should draw on cross-curricula knowledge in responding to real-world contexts. Further, they should incorporate opportunities to showcase transversal attributes such as digital literacy, collaboration, creativity and critical thinking (Paramore, 2017).

Assessment identity

Assessment literacy is a multi-dimensional construct that includes affective elements that shape the role of teachers in assessment (DeLuca et al., 2019). Teachers' feelings, emotions, values, and beliefs about assessment derive from their own personal experiences. Teachers' positive beliefs and attitudes about the role that assessment plays in improving learning, as well as confidence and belief in their assessment skills, are an important part of their identity (Gotch & McLean, 2019). Aligning teachers' assessment practices with the relevant legislative and cultural context is also of crucial importance (Birenbaum et al., 2015).

Professional development

Having support and encouragement from school leaders can greatly enhance teachers' professional development (Adie et al., 2020). Teacher assessment practices continue to be highlighted in many countries as an area in need of further professional development. In this paper, we analyse the assessment focus within the Australian Professional Standards for Teachers (APSTs). Teachers should feel free to try out innovative and experimental approaches to assessment in their professional practices, and to have access to the expertise needed to do so. Schools can create enabling structures for assessment literacy among teachers by ensuring alignment between teachers' assessment strategies and principles set out in official school policies (DeLuca et al., 2016). Continuous access to good quality professional learning opportunities will ensure that teachers employ best practices and approaches to assessment.

School environment

In supportive environments, schools create enabling structures for assessment literacy among teachers (DeLuca et al., 2016). Supportive environments are those in which school leaders clearly articulate the importance of using data to inform improvements (ACER, 2016). Data-informed decision-making is important for school improvement (Schildkamp et al., 2019) and one way to empower teachers to optimise data use in schools is by using data teams (Crone et al., 2016). These teams can help analyse assessment data, design interventions and support colleagues to use data in informing teaching practices.

Engagement of learners

In addition to developing their own assessment literacy, teachers should also support learners to gain assessment literacy. This means ensuring that learners understand success criteria and engage in self- and peer- assessment. Further, peer- and self-assessment are important for learners to become independent; teachers can facilitate this by modelling how to make judgements about the quality of work (Dixon et al., 2011). Using questioning, probing and other instructional practices can enhance the effectiveness of assessment by helping learners to identify their strengths and areas for improvement (Duckor & Holmberg, 2019).

Integration of digital assessment

Digital assessments can improve efficiency in marking, moderating, and storing information, enabling teachers to use their resources better (Oldfield et al., 2012). These assessments provide opportunities to assess complex knowledge and reasoning that may not be possible to assess through traditional, paper-based methods (Jamil et al., 2012). To make the most of digital assessment, teachers need to undergo training in digital assessment design and undertake digital assessments themselves (Walker, 2007). Greater exposure to digital assessment can also improve teachers' abilities to design digital assessments. Professional development needs to constantly evolve as digital technologies evolve.

Conclusion

Assessment literacy has clearly evolved significantly over time and continues to evolve. While much of the current work in assessment literacy focuses on using assessment, there is much less focus on designing assessment or interpreting its outcomes, and the kind of competencies that teachers need to have to succeed in designing, implementing and enhancing assessments (Koh & Chai, 2016).

Equally, while digital technologies create new opportunities for assessment, very little attention has been paid to how teachers can make the most of digital tools in their assessment practices (Whitelock, 2011). More research is needed to fully evaluate how the changing nature of assessments is impacting teachers' assessment literacy.

This paper is based on:

Girgla, A., Good, L., Krstic, S., McGinley, B., Richardson, S., Sneider-Gregory, S., & Star, J. (2021). *Developing a teachers' assessment literacy and design competence framework*. <https://www.ibo.org/research/curriculum-research/cross-programme/developing-a-teachers-assessment-literacy-and-design-competence-framework-2021/>

References

- ACER (2016). *The National School Improvement Tool*. <https://www.acer.org/au/school-improvement/improvement-tools/national-school-improvement-tool>
- ACER. (2021). *Developing a teachers' assessment literacy and design competence framework*. International Baccalaureate Organization.
- Adie, L., Stobart, G., & Cumming, J. (2020). The construction of the teacher as expert assessor. *Asia-Pacific Journal of Teacher Education*, 48(4), 436–453.
- Birenbaum, M., DeLuca, C., Earl, L., Heritage, M., Klenowski, V., Looney, A., Smith, K., Timperley, H., Volante, L., & Wyatt-Smith, C. (2015). International trends in the implementation of assessment for learning: Implications for policy and practice. *Policy Futures in Education*, 13(1), 117–140.
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30(1), 3–12.
- Choi, K. M., Hwang, J., Jensen, J., & Hong, D. S. (2021). Teachers' use of assessment data for instructional decision making. *International Journal of Mathematical Education in Science and Technology*, 1–8.

- Coombs, A., DeLuca, C., LaPointe-McEwan, D., & Chalas, A. (2018). Changing approaches to classroom assessment: An empirical study across teacher career stages. *Teaching and Teacher Education, 71*, 134–144.
- Crichton, H., & McDaid, A. (2016). Learning intentions and success criteria: learners' and teachers' views. *The Curriculum Journal, 27*(2), 190-203.
- Crone, D., Carlson, S., Haack, M., Kennedy, P., Baker, S., & Fien, H. (2016). Data-based decision-making teams in middle school: Observations and implications from the middle school intervention project. *Assessment for Effective Intervention, 41*(2): 79–93.
- DeLuca, C., Chapman-Chin, A., & Klinger, D. A. (2019). Toward a teacher professional learning continuum in assessment for learning. *Educational Assessment, 24*(4), 267–285.
- DeLuca, C., LaPointe-McEwan, D., & Luhanga, U. (2016). Approaches to classroom assessment inventory: A new instrument to support teacher assessment literacy. *Educational Assessment, 21*(4), 248–266.
- DeLuca, C., & Bellara, A. (2013). The current state of assessment education. *Journal of Teacher Education, 64*(4), 356–372.
- Dixon, H. R., Hawe, E., & Parr, J. (2011). Enacting assessment for learning: The beliefs practice nexus. *Assessment in Education: Principles, Policy & Practice, 18*(4), 365–379.
- Duckor, B., & Holmberg, C. (2019). Exploring how to model formative assessment trajectories of posing-pausing-probing practices: Toward a teacher learning progressions framework for the study of novice teachers. *Journal of Educational Measurement, 56*(4), 836–890.
- Gotch, C. M., & McLean, C. (2019). Teacher outcomes from a statewide initiative to build assessment literacy. *Studies in Educational Evaluation, 62*, 30–36.
- Hattie, J. (2009). *Visible Learning: A Synthesis of Over 800 Meta-analyses Relating to Achievement*. Routledge.
- Jamil, M., Tariq, R. H., & Shami, P. A. (2012). Computer-based vs paper-based examinations: Perceptions of university teachers. *Turkish Online Journal of Educational Technology - TOJET, 11*(4), 371–381.
- Koh, J. H. L., & Chai, C. S. (2016). Seven design frames that teachers use when considering technological pedagogical content knowledge (TPACK). *Computers & Education, 102*, 244–257.
- Kyriacou, C. (2010). *Effective teaching in schools: Theory and practice*. 3rd ed. Oxford University Press.
- Looney, A., Cumming, J., van Der Kleij, F., & Harris, K. (2018). Reconceptualising the role of teachers as assessors: Teacher assessment identity. *Assessment in Education: Principles, Policy & Practice, 25*(5), 442–467.
- Mertler, C. A. (2009). Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving schools, 12*(2), 101-113.
- Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment in every classroom: A guide for instructional leaders*. ACSD.
- Oldfield, A., Broadfoot, P., Sutherland, R., & Timmis, S. (2012). *Assessment in a digital age: A research review*. University of Bristol.
- Paramore, J. (2017). Questioning to stimulate dialogue. In R. Paige, S. Lambert S. & R. Geeson (Eds.). *Building skills for effective primary teaching* (pp. 125–142). Learning Matters.

- Pastore, S., & Andrade, H. L. (2019). Teacher assessment literacy: A three-dimensional model. *Teaching and Teacher Education, 84*, 128–138
- Perera, H., Calkins, C., & Part, R. (2019). Teacher self-efficacy profiles: Determinants, outcomes, and generalizability across teaching level, *Contemporary Educational Psychology, 58*: 186–203.
- Schildkamp, K., Poortman, C., Ebbeler, J. & Pieters, J. (2019). How school leaders can build effective data teams: Five building blocks for a new wave of data-informed decision making. *Journal of Educational Change, 20*, 283–325.
- Walker, D. J. (2007, May). Principles of good online assessment design. In *International Online Conference on Assessment Design for Learner Responsibility. Re-Engineering Assessment Practices in Scottish Higher Education (REAP)*.
- Whitelock, D. (2011). Activating assessment for learning: are we on the way with web 2.0? *Web 2.0-based e-learning: Applying social informatics for tertiary teaching* (pp. 319-342). IGI Global.
- Xu, H. (2017). Exploring novice EFL teachers' classroom assessment literacy development: A three-year longitudinal study. *The Asia-Pacific Education Researcher, 26*(3), 219–226.
- Yan, Z., & Cheng, E. C. K. (2015). Primary teachers' attitudes, intentions and practices regarding formative assessment. *Teaching and Teacher Education, 45*, 128–136.

CONFERENCE PROGRAM

Day 1: Monday 22 August

Keynote

Reimagining the purpose of assessment

11:15am – 12:15pm AEST

Professor Geoff Masters AO, ACER

ACER models of writing: Changing the assessment mindset

12:30pm – 1:30pm AEST

Juliette Mendelovits and Judy Nixon, ACER

Reimagining classroom assessment and feedback to meet learner needs

1:45pm – 2:45pm AEST

Dr Fabienne Van der Kleij, ACER

Break

Twilight session: Assessments for global goals

7:00pm – 8:00pm AEST

Dr Ursula Schwantner, Dr Sally Robertson, Dr Dara Ramalingam, Dr Tim Friedman, ACER

CONFERENCE PROGRAM

Day 2: Tuesday 23 August

Reimagining assessment in culturally responsive ways

11:15am – 12:15pm AEST

Dr Carly Steele and Associate Professor Graeme Gower, Curtin University

Using PAT data to inform teaching and learning

12:30pm – 1:30pm AEST

Marc Kralj, Dr Shani Sniedze-Gregory, Rachel Felgate and Caithlin Power, ACER, and Grant Barclay and Darren Leech from Marist College, Canberra

Keynote

An innovative method for teachers to formatively assess writing online

1:45pm – 2:45pm AEST

Dr Sandy Heldsinger, Brightpath and Associate Professor Stephen Humphry, The University of Western Australia (UWA)

Break

Twilight session: Sharing and securing learners' performance standards across schools

7:00pm – 8:00pm AEST

Emeritus Professor Richard Kimbell, Goldsmiths, University of London

CONFERENCE PROGRAM

Day 3: Wednesday 24 August

Playful assessment: Don't stop the fun

11:15am – 12:15pm AEST

Louisa Rosenheck, Kahoot! and Dr YJ Kim, University of Wisconsin-Madison

The transformational change of adaptive assessment

12:30pm – 1:30pm AEST

Dr Jarrod Hingston and Toby Newton, ACER

Keynote

Assessment moderation: Is it fit for purpose?

1:45pm – 2:45pm AEST

Associate Professor Lenore Adie, Australian Catholic University (ACU)

Graduation Ceremony

2022 Graduate Certificate in Education: Assessment of Student Learning

3:00pm – 4:00pm AEST

Break

Twilight session: The accuracy of item and ability estimate of concurrent calibration under the Interlaced Anchor Test design administration

7:00pm – 8:00pm AEST

Dr Urip Purwono and Dr Bakir Haryanto, ACER Indonesia

CONFERENCE PROGRAM

Day 4: Thursday 25 August

Keynote

Karmel Oration

Making learning visible: Moving from nouns to verbs

11:15am – 12:15pm AEST

Dr Diane DeBacker, Center for Certification and Competency Based Education, University of Kansas

Assessment in the early years: A symposium about measurement, applications and going to scale

12:30pm – 1:30pm AEST

Dr Dan Cloney, ACER, and invited guest speakers: Kellie Picker, Dave Jeffries, ACER; Sarah Groom, Mary-Ruth Mendel, Eric Brace, Australian Literacy and Numeracy Foundation; Myra Geddes, Lisa Palethorpe, Goodstart Early Learning

Assessing reading: How assessment can be used to target teaching and enhance understanding of reading comprehension

1:45pm – 2:45pm AEST

Dr Sandra Knowles, ACER

Break

Twilight session: Teachers' assessment literacy and design competence framework

7:00pm – 8:00pm AEST

Dr Sladana Krstic and Dr Sarah Richardson (ACER) with Dr Sarah Manlove, the International Baccalaureate Organization

Conference close

8:00pm AEST

CONFERENCE PROGRAM

Masterclass: Friday 26 August

Masterclass 1: Using assessment for student self-regulated learning

9:00am – 1:00pm AEST

Dr Katie Richardson, Dr Fabienne van der Kleij, Dr Claire Scoular, Dr Ian Teo and Dr Sandra Knowles, ACER

Masterclass: Monday 29 August

Masterclass 2: ACER's Learning Progressions

9:00am – 1:00pm AEST

Charlotte Waters, Prue Anderson, Ross Turner, Dr Sandra Knowles, Dr Dara Ramalingam, Stavroula Zoumboulis and Kat McGill, ACER