Research Conference 2021

Excellent progress for every student

Proceedings and Program

ONLINE EVENT

16–20 August 2021

Australian Council *for* Educational Research

ACER

Program and Organising Committee

Dr Claire Scoular
Conference Convenor

Professor Geoff Masters AO, CEO, ACER
Professor Ray Adams
Daniela Buehrig
Alex Harkness
Gareth Lloyd
Julia Mattea
Dr Catherine McClellan
Emma Noble
Kelly Van Os
Ralph Saubern
Ross Turner
Charlotte Waters
Kelly Williams

ISBN 978-1-74286-638-3
https://doi.org/10.37517/978-1-74286-638-3

19 Prospect Hill Road
Camberwell VIC 3124
Phone: +61 3 9277 5555
www.acer.org

Design by ACER Creative Services
Edited by Kylie Cockle

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 Indigenous learning, education and research.

CONTENTS

FOREWORD

Excellent progress for every student

Welcome to Research Conference 2021.

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 evidence-based strategies for ensuring that every student makes excellent ongoing progress in their learning. This is an important topic because many students in our schools do not make good, steady progress. Some slip behind and fall increasingly behind the longer they are in school. By the middle years of school, this contributes to significant levels of student disengagement.

ACER has invited a number of leading educational researchers to join us to share the findings of their research relevant to this topic. An important conclusion of this research is that support for learners' ongoing progress depends on a deep understanding of the nature of progress itself – that is, an understanding of how more sophisticated knowledge in an area of learning, deeper understandings, and higher levels of skill typically unfold over extended periods of time, often across many years of school. When the nature of long-term progress is well understood and documented, it provides a frame of reference for establishing the points individuals have reached in their learning, identifying useful next steps for teaching, and monitoring learning growth over time.

For the first time in its 25-year history, the Research Conference is entirely online. There are obvious disadvantages in not being able to meet in person, but an advantage is that we have been able to include international speakers who would not have been able to travel. This has enhanced our ability to feature the latest research. It also means that we have conference delegates joining us virtually from multiple locations, and that we have been able to distribute presentations over a number of days.

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

Professor Geoff Masters AO
CEO, Australian Council for Educational Research

KEYNOTE

How education gets in the way of learning

Professor Geoff Masters AO
Australian Council for Educational Research

https://doi.org/10.37517/978-1-74286-638-3_1

Geoff Masters is the CEO and board member of the Australian Council for Educational Research. He has a PhD in measurement, evaluation and statistical analysis from the University of Chicago and has published widely in the fields of educational measurement, educational assessment and school improvement. Geoff has conducted a number of reviews for governments, including a review of examination procedures in the NSW Higher School Certificate; an investigation of options for the introduction of an Australian Certificate of Education; a national review of options for reporting and comparing school performances; reviews of strategies for improving literacy and numeracy learning in government schools in Queensland and the Northern Territory; and a review of senior secondary assessment and tertiary entrance procedures in Queensland. Most recently, he has reviewed the New South Wales school curriculum. Geoff's 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

The formal structures and processes of school education – including the organisation of the school curriculum, processes for assessing student learning, methods of reporting performance, and the uses to which student results are put – are often inconsistent with what is now known about the best ways to promote human learning. Rather than being designed to maximise every student's learning, these structures and processes often reflect 20th century priorities, including the use of school education to sort and select students into different education and training destinations, and future careers. This sorting function of schooling is becoming increasingly irrelevant in knowledge economies that now look to their school systems to provide every student with high levels of knowledge, understanding and skill, including skills in critical and creative thinking, problem-solving, using new technologies, and working collaboratively with others. The challenge is to ensure that every student reaches the levels currently achieved by only some. However, the structures and processes of today's schools are often poorly designed to meet this challenge.

The curriculum

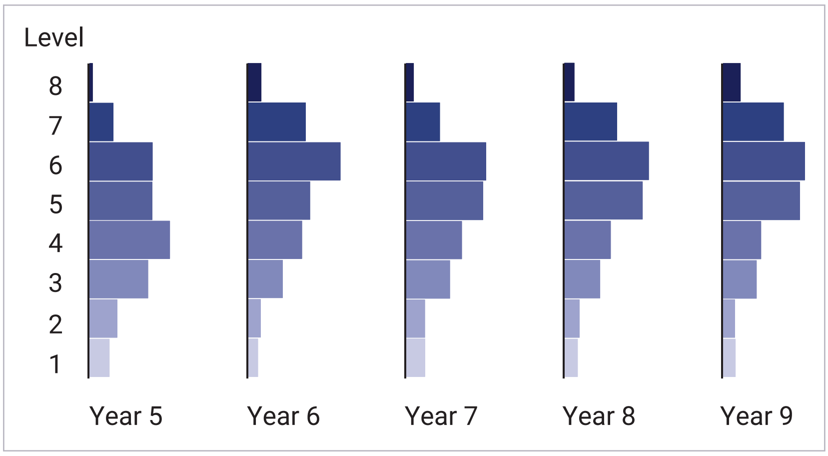
Consider, for example, the school curriculum. In its design, the current curriculum has much in common with an industrial-era assembly line. All students move along it at the same rate. Each year, the same curriculum is delivered to all students who are given the same amount of time to master it. They then move in lockstep to the next year's curriculum where the process is repeated. Students who have not mastered the content of the current year's curriculum and lack the prerequisites for the following year's curriculum move on regardless. Other students, who may not have required a full year to do this, are unable to advance to a more challenging curriculum until everybody moves in unison.

The consequence of this design is predictable. Students who lack the prerequisites for the next year's curriculum are also less likely to master that curriculum. As a result, they may be even less likely to master the curriculum in the year after that, and so on. Not surprisingly, many students fall increasingly far behind the longer they are in school and as year-level curricula become increasingly beyond their reach. By the middle years of school, many students have fallen so far behind that they begin to disengage. Some stop attending school entirely.

The curriculum designed in this way is an excellent sorting mechanism. It sorts out struggling students. Its effectiveness in doing this is clear from data on student achievement. Consider, for example, the data in Figure 1. These were collected by Professor Di Siemon and her colleagues who began by defining a number of levels of mathematical knowledge. These levels were not based on how old students were or what grade of school they were in; they were simply levels of increasing mathematical knowledge and understanding. Level 1 was the lowest, Level 8, the highest.

They then assessed Australia-wide samples of students in each of Years 5 to 9 to see which levels of mathematical knowledge students demonstrated. Figure 1 shows the results. At the top left, it can be seen that some Year 5 students were already at the highest level of mathematical knowledge – Level 8. At the bottom right, some Year 9 students were still at the lowest level – Level 1. Siemon and her colleagues concluded that the spread within each year level was equivalent to about seven years of mathematics learning. Students toward the bottom right in Figure 1 no doubt have struggled with year-level mathematics curricula throughout their schooling. The Year 9 mathematics curriculum is now well beyond their reach and they have effectively been sorted out of the system.

Figure 1 Levels of mathematics knowledge – Australian students, Years 5–9



Note: The spread within each year level represents a range in students' mathematics achievement equivalent to seven years of schooling
Source: Adapted from Siemon (2019, p. 13)[1]

In contrast to the current delivery of the curriculum, research shows that the way to maximise an individual's learning is to provide them with learning opportunities at an appropriate level of (stretch) challenge. People do not learn effectively when presented with material for which they lack prerequisite knowledge or skills or when given material well within their comfort zones. Learning opportunities need to be well targeted to individuals' present levels of attainment and readiness. For many students, this is precisely what the time-driven assembly-line curriculum, with its assumption that all students in the same year of school are more or less equally ready for the same learning, fails to do.

The school curriculum is also reminiscent of industrial processes in that it commonly breaks down subjects into component elements, usually referred to as 'outcomes' or 'objectives'. The intention is that each outcome will be taught and mastered. Many outcomes of this kind can result in a 'crowded'

1 Siemon, D. (2019). Knowing and building on what students know: The case of multiplicative thinking. In D. Siemon, T. Barkatsas, & R. Seah (Eds.), *Researching and using progressions (trajectories) in mathematics education*. BRILL. https://doi.org/10.1163/9789004396449

curriculum that teachers feel required to cover in the limited time available. Teachers also may be led to believe that they should assess every outcome individually. Under this thinking, the curriculum becomes a checklist of outcomes, all of which may appear equally important, with curriculum coverage taking precedence over depth of learning.

In contrast, research into learning underscores the importance of learners developing deep understandings of a relatively small number of essential concepts, principles and methods in an area of learning. These essential concepts, principles and methods provide a frame of reference for organising and understanding specific facts and procedures. The development of deep understanding usually requires significant time and may occur only over multiple years of schooling. In general, deep understanding and thinking are not 'outcomes' that can be recorded as present or absent on a checklist; they are reflected in a learner's ability to transfer and apply their knowledge to a growing range of unseen and increasingly complex situations.

Assessing and reporting

If the curriculum is presented as an assembly line, with all students being delivered the same content at the same time and being given the same amount of time to master it, then the important question is how much of that content each student has learnt and can demonstrate. At the end of each year, students are assessed and graded on that year's content before moving to the next year's curriculum. The fundamental purpose is to determine how well students have learnt what has been taught. Assessments also may be undertaken during the year to identify individual gaps in learning, provide feedback, and identify material that may need to be retaught, but whenever assessment occurs, its core purpose under this model is to judge how well students have learnt a body of taught content. And these judgements typically are conveyed as marks, percentages or grades.

A consequence of this approach is that each new school year is treated as a fresh start, largely independent of the prior year. The intention is that the grade a student receives on the year's curriculum should be uninfluenced by the grade they received on the previous year's curriculum; years of school are independent stations on the assembly line. As a result, a student can receive the same grade every year. A student who receives a grade of, say D, year after year is given little sense of the absolute progress they are making and, worse, may be sent a message that there is something stable about their ability to learn (they are a D-student).

Again, in industrial (and agricultural) settings, it is common practice to grade the products or outcomes of production processes for their quality. The assessment and grading of student performance on each year-level curriculum mirrors this practice.

In contrast, research into learning reveals the importance of seeing learning as a continuous process through which learners build new learning on prior learning. Learning is less a process of mastering isolated outcomes on a checklist, and more an ongoing process of developing increasingly sophisticated knowledge, deeper understandings, and higher levels of skill over time. One of the most effective ways to promote learning is to establish the point a learner has reached in their progress and to provide well-targeted learning opportunities that build from that point. This is the realisation that underpinned Lev Vygotsky's Zone of Proximal Development[2] and David Ausubel's observation more than half a century ago:

> If I have to reduce all of educational psychology to just one principle, I would say this: the most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly (Ausubel, 1968, p. vi).[3]

2 Vygotsky, L. S. (1978). *Mind in society.* Harvard University Press.
3 Ausubel, D. P. (1968). *Educational psychology: A cognitive view.* Holt, Rinehart and Winston Inc.

Two implications follow from these research findings. First, all learning builds on prior learning and lays the foundations for potential future learning. Learning occurs when connections are built between new information and existing knowledge. For these reasons, learning is best conceptualised as a continuous process, and the sequencing of learning opportunities to build increasing depth is a key to successful teaching and learning. Second, to build from the point an individual has reached in their learning, it is necessary to 'ascertain' that point. Thus a focus on learning gives assessment a new purpose. No longer is the purpose to determine how well students have learnt what they have been taught; the new purpose is to establish where learners are in their long-term learning progress. At a fundamental level, the assessment purpose switches from judging to understanding.

In practice, the education process also sometimes includes selection points at which only some students are chosen for entry to the next education phase or course. At these times, the main purpose of assessment is not to inform learning, or even to grade performance, but to compare and select individuals. Because of the consequences for students, such assessments often distort teaching and learning. For example, if written examinations are a major component of the selection process, teaching and learning may overemphasise the recall of knowledge and theory and under-emphasise the ability to apply knowledge, critical and creative thinking, and problem-solving. Selection processes need to be recognised as external influences that have more in common with assembly lines than human learning.

When the focus is on learning rather than grading or selecting, student progress (or growth) becomes a critical consideration. As already noted, this concept is largely absent from an education model that prioritises the achievement of individual outcomes, grades year-level performances, and includes high-stakes assessments for managing student flows. Learning, on the other hand, leads to student progress/growth in the form of increasingly sophisticated knowledge, deeper understandings, and higher levels of skill. To draw inferences about learning, it is necessary to evaluate progress toward higher levels of knowledge, understanding and skill.

Under a learning model, assessments are designed to establish the points individuals have reached in their learning (what they know, understand and can do at the time of assessment), conceptualised as positions on a continuum or progression of learning. These guide next steps in teaching and learning. Such assessments also can be used to monitor learning progress over time – essential information for evaluating the effectiveness of teaching strategies and interventions and if students are to see and appreciate the progress they are making and to develop positive views of themselves as learners.

In summary, current education structures and processes, including the organisation and delivery of the curriculum, and assessment and reporting arrangements, are largely 20th century inventions introduced as a response to mass education. This response includes grouping all students by age, delivering all students the same content for the same amount of time, grading performances on this common content, and then advancing students in unison to the next stage of the curriculum. Australian schooling has not always been organised in this way, and is highly unlikely to be organised in this way in the future. The alternative will recognise that, despite efforts to standardise learning, individuals remain at very different points in their learning, make progress at different rates, and have very different learning needs. The future organisation of schooling will be guided more by our emerging understandings of human learning than by education models of the past.

Karmel Oration
Excellent progress for all: A function of year-level curriculum or evidenced-based learning progressions?

Professor Dianne Siemon
RMIT University

https://doi.org/10.37517/978-1-74286-638-3_4

Dianne Siemon is an Emeritus Professor at RMIT University. From 2014 to 2018, Di led the Reframing Mathematical Futures II research team that worked with 32 secondary schools nationally to develop evidenced-based learning progressions for algebraic, geometrical and statistical reasoning in the middle years. The project also produced validated assessment tools and targeted teaching advice to help identify and scaffold students' learning in relation to the progressions.

Di has been involved in pre-service mathematics education for over 30 years and remains actively involved in the professional development of practising teachers, particularly in relation to the development of the 'big ideas' in Number and the use of rich assessment tasks to inform teaching. Di has directed a number of other large-scale research projects including Building Community Capital to Support Sustainable Numeracy Education in Remote Locations (2006–2009), Scaffolding Numeracy in the Middle Years (2003–2006), Researching Numeracy Teaching Approaches in Primary Schools (2001–2003), and the Middle Years Numeracy Research (1999–2001). Di is a life member of the Australian Association of Mathematics Teachers and the Mathematical Association of Victoria.

Abstract

Excellent progress for all students is an ambitious but necessary goal if we are to improve the life choices of all students. At the moment, we are not serving all our students well despite the best efforts of teachers. We need to look further afield to the curriculum and assessment regimes that drive current practice. Grouping students by ability and offering a watered-down curriculum for some is not the answer. Evidenced-based learning progressions that point to what is important in ensuring all students build a deep, well-connected understanding of mathematics over time is what is needed to support reform at scale. Where the evidenced-based tools and resources produced by this type of research are used to identify and respond to student learning needs in relation to what is important, it has been shown to make a significant difference to student outcomes and engagement. Adopting a targeted teaching approach means that not everything has to be differentiated and not everything needs to be considered as often or to the same depth. Time can be spent researching challenging but accessible tasks and developing a culture that supports and rewards persistence, effort and a growth mind-set.

Introduction

Excellent progress for every student is the ultimate goal of all those involved in education. However, the continuing decline in Australian students' performance on international assessments of mathematical literacy (Thomson et al., 2019, 2020), and concern about the continuing decline in the number of students able and willing to undertake the more advanced mathematics subjects in the final years of secondary school (Wienk, 2020), suggest that while some students are doing very well, schools are not supporting *all students* to make excellent progress in mathematics.

There is 'long tail' of students who may be up to seven years behind their higher achieving peers in curriculum terms (Masters, 2013; Siemon & Virgona, 2001), and who disproportionally come from disadvantaged or lower socio-economic backgrounds (Thomson, 2021). This is inequitable in a country that prides itself on a fair go for all and aspires to 'excellent progress for all students'.

There are many reasons for this state of affairs, but an important one is the expectation that all students, irrespective of where they are in their particular learning journey, should attempt to master all aspects of the mathematics curriculum at their year level. This situation is exacerbated by the fact that school mathematics tends to be presented as a set of disconnected skills to be demonstrated and practised rather than explored and discussed, with little or no indication of what aspects of mathematics are more important than others in progressing students' mathematics learning (Siemon et al., 2012).

This raises the important question of what constitutes excellent progress. One view of what this might involve is given in the Gonski-led *Review to achieve educational excellence in Australian schools* – that we should aspire to 'Deliver at least one year's growth in learning for every student every year' (Department of Education and Training, 2018, p. x). But my question is, 'growth in relation to what?' Is it in relation to all aspects of the year-level curriculum, or what learning progression research suggests are the key ideas needed to sustain and support further learning in mathematics?

If a *years' growth* is defined in terms of year-level curriculum, it serves to maintain the status quo. For example, a Year 4 student who is five or so years behind their high-achieving peers, will still be five or more years behind their high achieving peers in Year 8 even if they have all achieved at least one year's growth in all of the intervening years. This is not excellent progress. Accepting the status quo or considering alternatives to 'presenting the curriculum in year-level packages' (Masters, 2017, p. 2) risks entrenching the practice of ability grouping, which is known to have a negative impact on student learning, self-esteem and confidence, and reproduce social inequality (e.g. Boaler & Staples, 2008; Francis, 2019).

Another problem with the Gonski view is that 'at least' suggests something a bit more than one year's growth rather than a lot more. We can and should be aspiring to deliver whatever it takes to ensure all students have the opportunity to engage productively with the curriculum at their year level or beyond. Teaching all aspects of the curriculum and grouping by ability is not the answer. Identifying where students are in relation to important mathematics and focusing on what is known to make a difference through targeted teaching and creative mixed ability teaching is what is needed (e.g. Breed, 2011; Goss et al., 2015; Siemon, Banks, & Prasad, 2018; Sullivan, 2011). And this is where research-based learning progressions/trajectories (LP/Ts) like the ones we have developed for multiplicative thinking (Siemon et al., 2006; Siemon, 2019) and mathematical reasoning (Siemon, Callingham, et al., 2018; Siemon et al., 2019) can contribute.

In recent years, attention has turned to the development of evidenced-based LP/Ts as a means of identifying what mathematics is important in ensuring students build a deep, well connected understanding of mathematics over time (e.g. Confrey et al., 2014; Siemon et al., 2006, 2019). Typically, this work also provides assessment tools to identify where learners are in their mathematics learning journey, and instructional materials and/or teaching advice aimed at progressing that journey from naïve to more sophisticated understandings.

What differentiates this work from the sort of progressions derived from the analysis of large data sets, such as NAPLAN, is that it is independent of the curriculum. The purpose of LP/T research is to identify and test ordered networks of key ideas and strategies rather than determine where students might be in relation to the curriculum. LP/T research starts out with an extensive review of the literature in a particular field to identify key ideas and establish a hypothetical learning trajectory. Assessment tasks are devised and trialled to test the hypothetical trajectory, and the results analysed using Rasch analysis (Bond & Fox, 2015), which allows item difficulty and student performance to be measured by the same unit and placed jointly on a single scale. This approach supports inferences about what students located at roughly the same points on the scale are able to do and what is likely to be within their grasp with teaching targeted to learning need, as detailed analyses of items located at similar points on the scale can be used to develop teaching advice about where to go to next.

In developing the LP/Ts for multiplicative thinking and mathematical reasoning, rich assessment tasks and partial credit scoring rubrics (Masters, 1982) were used to identify qualitative differences in students thinking, and to identify the 'big ideas' without which progress in these domains would be impacted. For multiplicative thinking, this work resulted in two assessment options that can be used to monitor students' growth over time on the same scale, and an evidenced-based LP/T that includes targeted teaching advice in the form of the *Learning assessment framework for multiplicative thinking* (LAF).[1] For mathematical reasoning, the research resulted in four assessment options for each area of mathematical reasoning (i.e. algebraic, geometrical, and statistical reasoning) and evidenced-based LP/Ts and targeted teaching advice for each area.[2]

Taken together these resources provide a basis for teachers 'to make decisions about the next steps in instruction that are likely to be better, or better founded than, the decisions they would have been made in the absence of that evidence' (Wiliam, 2011, p. 43). Understood and enacted in this way, research-based LP/Ts can sit alongside mandated curricula to help teachers identify where students are and where to go to next in relation to important mathematics. In addition, by pointing to what is critical in progressing students' mathematics learning *over time*, research-based LP/Ts can also be used to identify key stages in schooling by which certain big ideas need to be in place in order to support further mathematics learning. Prioritising these in planning is essential in preventing the mathematics achievement gap from widening the longer students spend in school. Targeted teaching has been found to progress students' mathematics learning by much more than one year's growth in less than a year (e.g. Breed, 2011; Siemon, 2017, Siemon et al., 2018a). It's not rocket science – focusing on those things that have been shown to make a difference, will make a difference.

Conclusion

Excellent progress for every student is an ambitious goal, but it will not be achieved by expecting all students in a particular year level to demonstrate every one of the content descriptors at that level at the same time in the same way. Neither is it an either/or situation as suggested by Gonski (Department of Education and Training. 2018). Evidenced-based LP/Ts that point to the 'big ideas' of mathematics, the connections between them, and their likely development over time, can help teachers better navigate the curriculum by pointing to what mathematics is important. This means that not everything needs to be differentiated and not everything needs to be considered as often or to the same depth. By providing valid assessment tools that identify where students are in relation to what is important, starting points for teaching can be identified; and by providing targeted teaching advice to support teachers to address student learning in relation to what is important, we can

1 https://www.education.vic.gov.au/school/teachers/teachingresources/discipline/maths/assessment/Pages/scaffoldnum.aspx
2 http://www.mathseducation.org.au/online-resources/introducing-the-rmfii-resources/

work towards excellent progress for all students. This will not be easy, but effective teaching never is. There is no single way to implement a targeted teaching approach and while some approaches appear to be better than others, successful implementation depends very much on the good will and commitment of all those involved, particularly school leadership.

References

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School, *Teachers College Record*, *110*(3), 608–645.

Bond, T., & Fox, C. (2015). Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.). Lawrence Erlbaum Associates

Breed, M. (2011). Constructing paths to multiplicative thinking [Unpublished doctoral dissertation]. RMIT University.

Confrey, J., Maloney, A., Nguyen, K., & Rupp, A. (2014). Equipartitioning, A foundation for rational number reasoning: Elucidation of a learning trajectory. In A. P. Maloney, J. Confrey, & K. H. Nguyen (Eds.), *Learning over time: Learning trajectories in mathematics education* (pp. 61–96). Information Age Publishing.

Department of Education and Training. (2018). *Through growth to achievement: Report of the Review to Achieve Educational Excellence in Australian Schools.* Commonwealth of Australia. https://www.dese.gov.au/quality-schools-package/resources/through-growth-achievement-report-review-achieve-educational-excellence-australian-schools

Francis, B., (2019). Re-assessing 'ability' grouping. Improving practice for equity and attainment. Taylor Francis Group.

Goss, P., Hunter, J., Romanes, D., & Parsonage, H. (2015). *Targeted teaching: How better use of data can improve student learning*. Grattan Institute.

Masters, G. (1982). A Rasch model for partial credit modelling. *Psychometrika*, *47*, 149–174. https://doi.org/10.1007/978-1-4757-2691-6_6

Masters, G. (2013). Reforming educational assessment: Imperatives, principles and challenges. *Australian Education Review*. https://research.acer.edu.au/cgi/viewcontent.cgi?article=1021&context=aer

Masters, G. (2017). *Lifting achievement levels and improving the return on Australia's investment in schooling. Submission to the Review to Achieve Educational Excellence in Australian Schools*. Australian Council for Educational Research. https://research.acer.edu.au/policy_analysis_misc/27

Siemon, D. (2017, 27 Feb.). Targeting the big ideas in mathematics. *Teacher*. https://www.teachermagazine.com/au_en/articles/targeting-big-ideas-in-mathematics

Siemon, D. (2019). Knowing and building on what students know: The case of multiplicative thinking. In D. Siemon, T. Barkatsas & R. Seah (Eds.), *Researching and using progressions (trajectories) in mathematics education* (pp. 6–31). Brill-Sense.

Siemon, D., Banks, N., & Prasad, S. (2018). Multiplicative thinking a STEM foundation. In T. Barkatsas, N. Carr, & G. Cooper (Ed.), *STEM education: An emerging field of inquiry*. Sense Publications.

Siemon, D., Barkatsas, T., & Seah, R. (Eds.) (2019). *Researching and using progressions (trajectories) in mathematics education*. Brill-Sense.

Siemon, D., Breed, M., Dole, S., Izard, J., & Virgona J. (2006). *Scaffolding numeracy in the middle years – Project findings, materials and resources, Final report*. https://www.education.vic.gov.au/school/teachers/teachingresources/discipline/maths/assessment/Pages/scaffoldnum.aspx

Siemon, D., Bleckly, J., & Neal, D. (2012). Working with the big ideas in number and the Australian Curriculum Mathematics. In W. Atweh, M. Goos, R Jorgensen & D. Siemon (Eds.) *Engaging the Australian Curriculum Mathematics – Perspectives from the field* (pp. 19–46). Mathematical Education Research Group of Australasia

Siemon, D., Callingham, R., Day, L., Horne, M., Seah, R., Stephens, M., & Watson, J. (2018). From research to practice: The case of mathematical reasoning. In J. Hunter, L. Darragh & P. Perger (Eds.), *Making waves, opening spaces, Proceedings of the 41st annual conference of the Mathematics Education Research Group of Australasia* (pp. 40–49). MERGA. https://merga.net.au/common/Uploaded%20files/Annual%20Conference%20Proceedings/2018%20CP%20Short%20communications/MERGA%20Book.pdf

Siemon, D., & Virgona, J. (2001). Roadmaps to numeracy – Reflections on the Middle Years Numeracy Research Project. *Proceedings of the Annual Conference of the Australian Association for Research in Education*, Fremantle.

Sullivan, P. (2011). *Teaching mathematics using research-informed strategies*. Australian Council for Educational Research.

Thomson, S. (2021). *Australia: PISA Australia – Excellence and equity?* ACER. https://research.acer.edu.au/ozpisa/51/

Thomson, S., De Bortoli, L., Underwood, C., & Schmid, M. (2019). PISA 2018: *Reporting Australia's results. Volume I student performance*. https://research.acer.edu.au/ozpisa/35/

Thomson, S., Wernert, N., Rodrigues, S., & O'Grady, E. (2020). *TIMSS 2019 Australia. Volume I: Student performance*. https://research.acer.edu.au/timss_2019/2/

Wienk, M. (2020). *The state of the mathematical sciences 2020*. AMSI. https://amsi.org.au/wp-content/uploads/2020/05/amsi-discipline-profile-2020.pdf

Wiliam, D. (2011). *Embedded formative assessment*. Solution Tree Press.

Accountable assessment

Professor Richard Lehrer
Vanderbilt University, Tennessee

https://doi.org/10.37517/978-1-74286-638-3_9

Richard Lehrer is Professor Emeritus and Research Professor of Education at Vanderbilt University. A former high school science teacher, he is a member of the National Academy of Education, a Fellow of the American Educational Research Association and a recipient of the American Psychological Association's Distinguished Contributions in Applications of Psychology to Education. Working with teachers, he focuses on the design of classroom learning environments that support the growth and development of children's understandings of how knowledge is generated and revised in mathematics and in sciences. His research examines productive means to engage children in forms of mathematics that are critical to STEM education, especially geometry, measure, data and chance. This research is aligned with parallel efforts in science education to induct children in the signature practice of the sciences–invention and revisions of models of natural systems. A contemporary collaboration with the Berkeley Evaluation and Assessment Research Center explores the feasibility of integrating teachers' classroom-embedded judgements of student learning into the kinds of psychometric models that are employed in standardised, 'accountability' assessments.

Abstract

There is widespread agreement about the importance of accounting for the extent to which educational systems advance student learning. Yet, the forms and formats of accountable assessments often ill serve students and teachers; the summative judgements of student performance that are typically employed to indicate proficiencies on benchmarks of student learning commonly fail to capture student performance in ways that are specific and actionable for teachers. Timing is another key barrier to the utility of summative assessment. In the US, summative evaluations occur at the end of the school year and may serve future students, but do not help teachers to better support the students who were tested. In contrast, formative assessments provide actionable grounds to improve the quality of instruction on the basis of both the granularity and specificity of their content and their timing. Unfortunately, the psychometric qualities of formative assessments are often unknown. I describe an innovative approach to assessment that aims to blend the productive characteristics of both summative and formative assessment. The resulting assessment system is accountable to students and teachers by providing actionable information for improving classroom instruction, and at the same time, it addresses the demands of psychometric quality for purposes of system accountability as it is currently practiced (in the US). The innovative assessment system relies on partnership with teachers to generate 1) a shared conceptual frame for describing instructional goals and valued forms of teaching and learning; 2) a set of electronic tools to help teachers detect, share, analyse, and interpret student learning data; and 3) classroom and school-level community professional development structures to support and sustain a widespread practice of assessing to guide instruction. These features are coupled with new psychometric models, developed by the Berkeley Evaluation and Assessment Research Center, that provide more robust estimates of student learning by linking information from multiple sources, including student classroom work, student responses to formative assessments, and summative evaluations. (Mark Wilson will address the psychometric modeling during this conference.) Here I describe challenges

and prospects for this innovation with a case study of its implementation in a K–5 elementary school that is seeking to improve the quality of instruction and students' understandings of measure and rational number arithmetic.

Introduction

Although the purposes of assessment are varied, there is widespread agreement about the importance of accounting for the extent to which educational systems advance student learning. Yet, the forms and formats of accountable assessments often ill serve students and teachers. In the US, summative evaluations used for accountability occur at the end of the school year. These evaluations could, in principle, serve future students, but they do not help teachers better support the students who were tested. Moreover, the implications of student performance on these summative evaluations for instruction tend to be very general, primarily because the tests are constructed in ways that are not well informed by constructs that describe typical progressions and patterns of student thinking (Wilson, 2005) . As a result, knowing that student performance in any area of mathematics is substandard does little to inform specific steps toward instructional improvement. In contrast, formative assessments are designed to provide actionable grounds to improve the quality of instruction due to increased granularity and specificity of their content and their timing (e.g. Black & Wiliam, 1998, 2009). As Wiliam (2015) clarifies, the signature of formative assessment is anticipating how students will think about situations posed during assessment and taking appropriate action accordingly. Unfortunately, the psychometric qualities of these forms of assessment are often unknown, and therefore are difficult to align with accountability assessments.

The premise of our collaboration with colleagues at University of California, Berkeley is that if ongoing assessment of student thinking is woven into the fabric of instruction, then teacher judgements of students' ways of thinking can inform psychometric modelling of student learning. Summative and ongoing formative assessments can be coordinated to generate more robust and actionable accounts of student learning. Moreover, assessment can be more accountable to the ongoing improvement of instructional practice and student learning in real time, rather than serving primarily as an aftermath to instruction. Achieving these goals means that teachers must learn to read and register selected forms of student thinking as they emerge during the course of classroom activity. Moreover, on the basis of what the data show, teachers must learn to leverage their knowledge of student thinking to improve the quality of instruction, so that assessment becomes a vital part of instructional practice.

Moreover, although most assessments are conducted by individual teachers, the practice of assessment, as well as its meaning and perceived value, are influenced by the surrounding community (Horn et al., 2015). In workgroups and grade teams, teachers communicate and subtly enforce a common epistemic orientation toward assessment (Horn et al., 2015). By epistemic orientation, Horn means teachers' perspectives – often tacit – on what can be known with data, how to know it, and why it is of value. Consequently, assessment practice is constituted by an interplay between individual teacher activity in a classroom or related instructional setting, and a teacher's anticipations of the norms and interpretations of the surrounding community. With this dual view of assessment practice in mind, we aimed to create and test an assessment system designed to address two coordinated purposes: 1) to provide ongoing, instructionally-productive evidence to teachers about student learning; and 2) to link dense information from student work products and formative assessments with summative assessments in new psychometric models that generate robust estimates of the growth of student learning. Such an assessment system includes:

- a conceptual frame shared by all participants for generating and interpreting evidence of learning in student activity across instructional settings

- a set of tools to amplify teachers' ability to detect, capture, share, analyse, and make sense of evidence of learning across instructional settings
- community structures across classroom, school and project partnership to support and sustain the practice of assessing to guide instruction.

To test the feasibility of this innovation, we collaborated with K–5 teachers in an intact school setting to construct an assessment system that would allow us, collectively, to track student learning of the mathematics of measure (length, angle, area, volume), and of children's learning of related concepts of rational number as teachers introduced measurement models to promote learning about fractional quantities and operators. The initial impetus for the focus on measure was children's comparatively poor prior performance on summative, statewide assessment in these areas of mathematics, as well as its many conceptual connections to a wide array of mathematical concepts taught in the elementary grades.

Constructing an assessment system

Participants

To construct the elements of an assessment system – a shared conceptual frame, appropriate tools, and productive community structures – we collaborated with 18 K–5 teachers, most of whom taught at Sleeve Elementary in the south-central region of the US. Three participating teachers were located at another school in the district. The district is the largest in the state. The student population of Sleeve Elementary is primarily rural and white. I met with teachers once each month for two to three days over two years (Summer 2018 – May, 2019; September 2019 – March 2020, interrupted by the suspension of schooling due to the COVID-19 pandemic). I also conducted multi-day summer institutes each year, once in person and once via Zoom conferencing. During the past year (August 2020– present), students attended school in person intermittently, and instruction was conducted online during the rest of the time. Access to digital instruction was especially problematic for many students.

Conceptual tools to promote shared vision

Supporting teachers to articulate a shared vision of instruction, learning and assessment included the design and iterative development (with teacher feedback and frequent contribution) of a set of conceptual tools. These included most prominently *constructs*, *lessons* and formative *assessment items* to support student learning of particular elements of constructs.

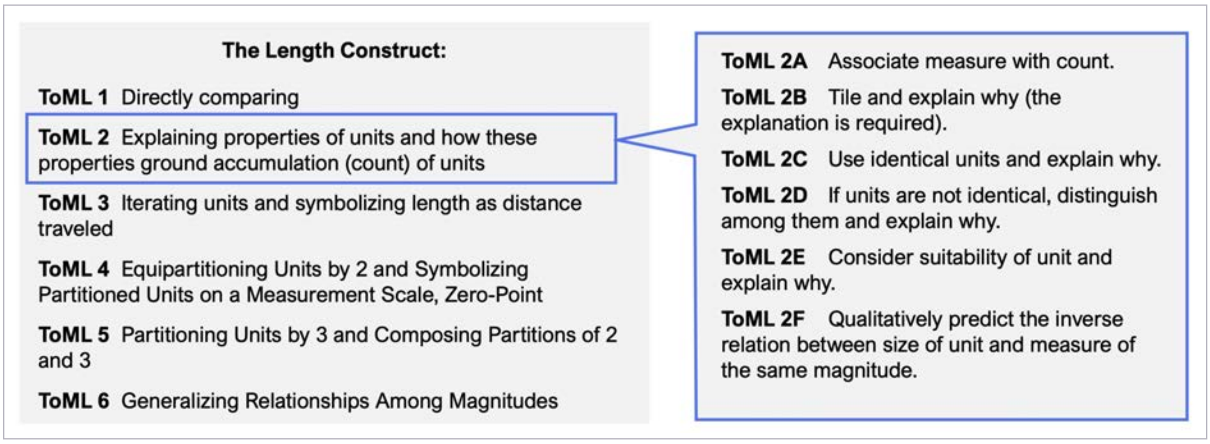
Constructs

Constructs identify typical forms of student thinking and articulate how these forms of thinking progress when they are appropriately supported by instruction (Wilson, 2005). The constructs are not fully-fledged theories of learning, but rather, are tuned to highlight aspects of learning that contribute to effective next instructional steps within specific content areas. Theories of learning are necessarily much finer-grained and more technical, and are not usually accurately described as linear (stage-like) paths through levels of a construct (Lehrer & Schauble, 2015).

Progress maps describe how children's thinking, as captured in constructs, usually develops. Progress maps are coarser-grained descriptions that are intelligible and practical; they represent an informational tradeoff for informing instruction. That is, they capture important variants in student thinking, but like all models, omit variations less commonly observed and forms of thinking that are not usually useful for guiding instruction. They set a local mathematical horizon that influences how teachers respond to students during the course of formative assessment. That is, they help teachers identify local 'next steps' in student thinking, so that they can decide upon reasonable approaches for supporting students' learning without having to manage a level of information that would otherwise be overwhelming (Kim & Lehrer, 2015).

We developed and refined four constructs that depict student progress in conceptions of the measure of length, angle, area, and volume. The constructs are organised as narratives of development and are summarised as tables of levels that describe and exemplify growth in students' ways of thinking. Each level is constituted by multiple sub-levels that collectively constitute the form of thinking characterised by that level. For example, initial levels (Levels 1 and 2) of the length construct specify how young children first begin to engage with the fundamental problematic of measure – identifying and characterising attributes to be measured and comparing values of these attributes directly and also indirectly via units of measure. Performances at these initial levels focus on properties of unit, such as the need to tile units without gaps or overlaps, and on understandings of the logical necessity that governs the performance (e.g. why gaps or overlaps produce inaccurate measures, not simply that they do). Figure 1 lists the levels of the length construct and, for Theory of Measure – Length (ToML) Level 2, illustrates how each level is composed of a network of related concepts that collectively are indexed by that level.

Figure 1 Theory of Measure – Length (ToML)



Lessons

Classroom lessons are designed to clarify how the conceptual change envisioned by constructs can be supported by instructional practices. For example, an image of length as dynamically generated by travelling from a starting point to a specified location often helps young children conceptualise length as a distance. This interpretation makes symbolisation of units on a ruler more intelligible, so that the location of 1 at an endpoint of a unit interval is interpreted as the distance travelled, rather than as merely marking one unit of a collection of units. Over the course of our collaboration, the lessons have undergone multiple rounds of revision and have been augmented with teacher-authored examples and alternatives represented in a 'teacher's corner'. Teachers and researchers regard lessons not as static structures, but subject to change as we collectively learn more about student thinking and how to support it.

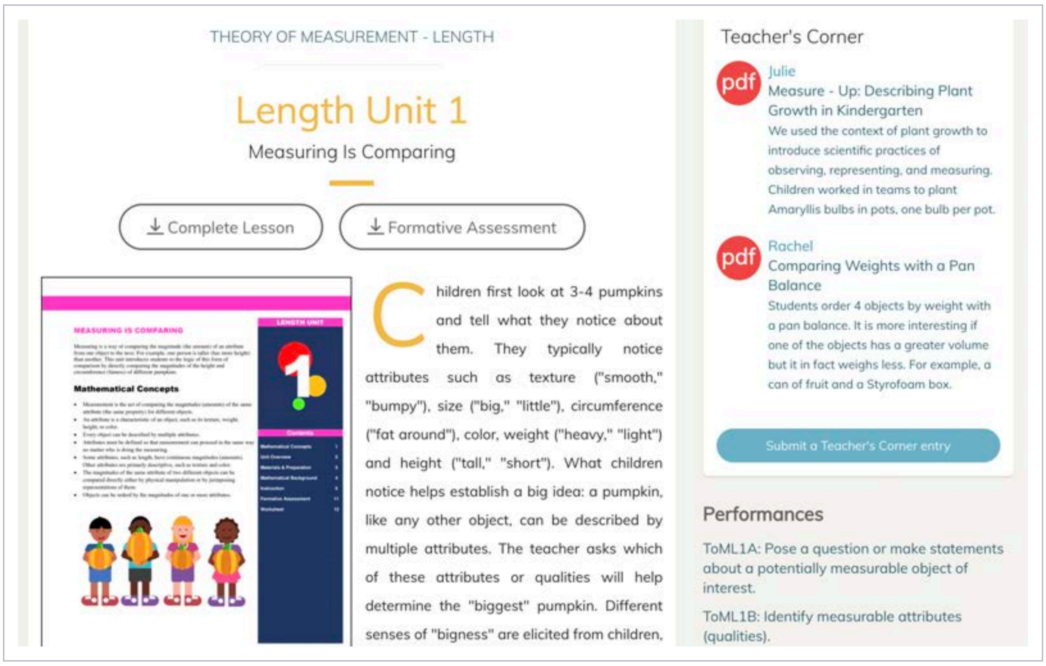
Formative assessment

Every lesson includes formative assessment items and illustrates how prospective student responses are aligned with particular levels on the construct map. For example, one of the formative assessment item displays six two-dimensional figures (including a line and a figure that is not closed) and asks students to circle all the figures that have an area. After students complete the formative assessment, and after the teacher has aligned student responses to levels of the relevant construct (in this case, the area construct), the teacher conducts a *formative assessment conversation* in which they juxtapose student responses and students explain the thinking that guided their responses.

In a follow-up discussion about the item just described, some Grade 3 students (7–8 years old) argued that it is possible to find an area measure for figures that are 'almost' closed. Rather than rejecting this proposal, the teacher asked children to justify their choices. At the board, students demonstrated how they would tile the entire space into which the area 'leaked'. Other students agreed that they could obtain a measure in this way, but objected that it would be difficult to know when to stop. Should one 'go to the road' outside the school? The teacher then drew 'large' and 'small' open figures, asking children to estimate the area measurement of each. Children concluded that all open figures would have the same (infinite) measure and conceded that this result would defeat the original intent to use measure to compare areas. Thus, rather than resorting to pre-determined definitions, the teacher supported students in reaching the consensus that it made most sense to restrict area measure to closed figures.

Constructs, lessons (including teacher elaborations), and formative assessments are available digitally, as illustrated in Figure 2. The district has adopted many of the lessons to guide their mathematics education program, although that also has had the unfortunate consequence that lessons have been incorporated into pacing guides and related forms of curricular control.

Figure 2 A suite of conceptual tools: lessons, constructs and formative assessments

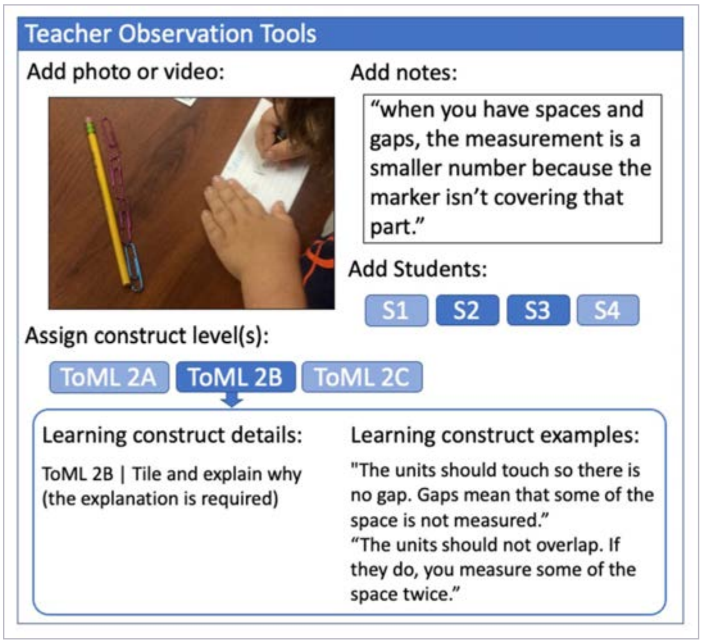


Digital tools to support ongoing assessment (designed by Corey Brady)

Teacher observation tools (TOTs)

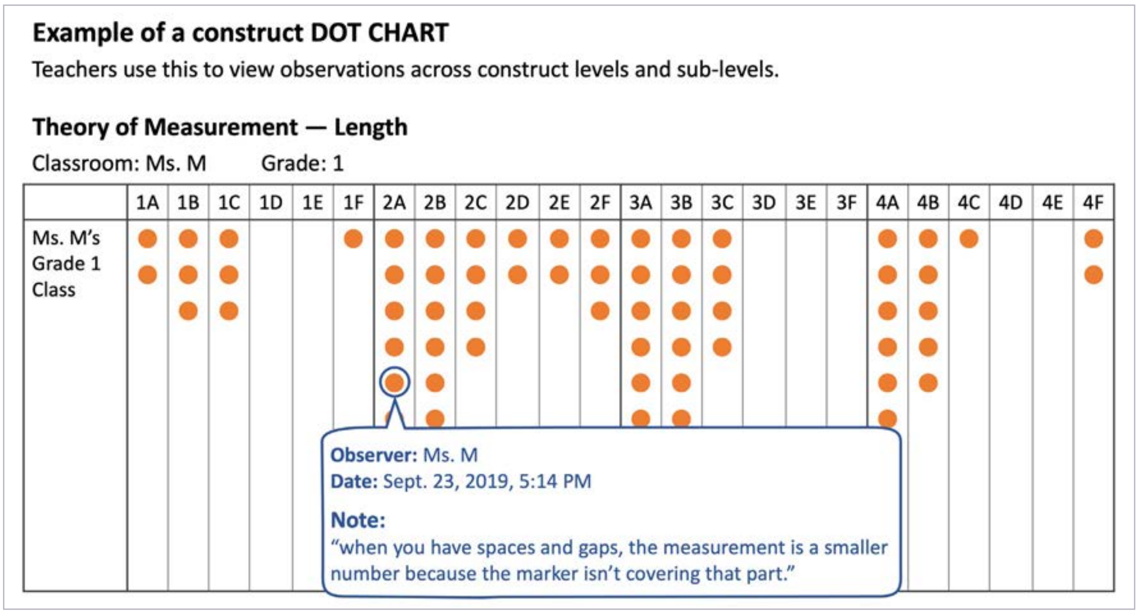
Teachers' judgements of students' ways of thinking are recorded with a web-based toolkit implemented on iPads. The toolkit allows teachers to record and store evidence of student thinking (typically video, photo, and field notes) that they observe during the course of instruction, and to associate this evidence with particular sub-levels of one or more constructs by means of a built-in coding system. This capacity extends the meaning of 'item' to include diverse expressions of student thinking as revealed by student talk, activity, and work products. Figure 3 is a facsimile of the recording portion of the toolkit. It exemplifies a photo and teacher note, the teacher's selection of the appropriate construct sub-level that describes one or more students' thinking, and attribution to one or more students.

Figure 3 Recording evidence of student thinking



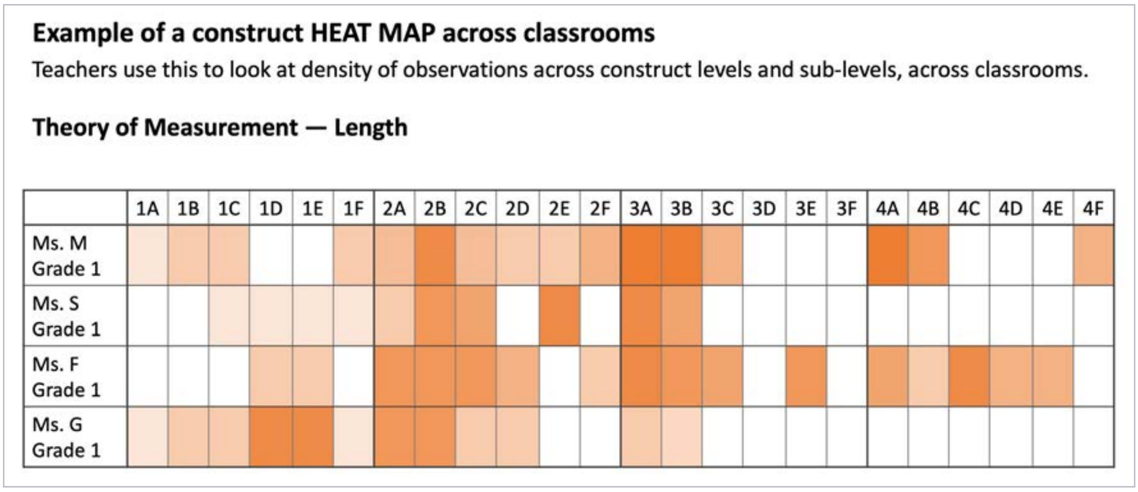
TOTs includes visualisations of student data that serve several functions – some for individual classroom teachers and others at a community-wide level. Figure 4 displays a facsimile of a dot plot of evidence for a construct from one teacher's classroom. Each dot corresponds to an observation and when selected the contents of the observation are revealed (here a portion of the previous observation is displayed). This display is handy for tracking evidence at the construct level for the class and provides a general picture of the class's current progress with respect to the given construct.

Figure 4 Dot plot of observations by construct sub-level.



A more economical display of data like these that seems to be preferred by teachers is a 'heat map' (see Figure 5), which uses color intensity represent frequency of observation. This view can also be used to represent observations across classrooms. This school-wide view is an important component of an emerging assessment practice in the school that is described in the next section.

Figure 5 Heat map of observations by construct sub-level across classrooms.



A 'star chart' view, depicted in Figure 6, represents observations at particular sub-levels of a construct for individual students, a feature that helps teachers ensure that their estimates of student learning are based on a census of students, and not a select few.

Figure 6 Student level evidence of learning

Example of a student STAR CHART

Teachers use this to follow observations of individual students across a construct of theory of measurement.

Theory of Measurement — Length

Classroom: Ms. M Grade: 1

	1A	1B	1C	1D	1E	1F	2A	2B	2C	2D	2E	2F	3A	3B	3C	3D	3E	3F	4A	4B	4C	4D	4E	4F
Student 1												★	★	★	★									
Student 2														★					★					
Student 3		★	★				★		★											★				
Student 4							★					★	★	★					★	★				★
Student 5							★	★			★			★					★	★				★

Observer: Ms. M
Date: Sept. 23, 2019, 5:14 PM

Note:
"when you have spaces and gaps, the measurement is a smaller number because the marker isn't covering that part."

Establishing a community of assessment practice

We collaborated with teachers to establish practices of assessment that were supported by the conceptual tools of constructs, lessons and formative assessments, and by the use of TOTs to generate evidence of student learning. As noted previously, our emphasis on community was informed by its critical role in the development of the professional discourse necessary for the improvement of instruction (e.g. Ball & Cohen, 1999, Desimone, 2009; Gibbons & Cobb, 2017) and by its critical role in generating productive norms for assessment (Horn et al., 2015). We faced several challenges in realising a collective vision. For many teachers, these forms of mathematics were not familiar, primarily because past instructional practice in the school had emphasised procedural competence with tools, such as protractors and rulers. A related challenge was that instructional practices did not include a repertoire of ways of helping students conceptualise measure. Instead, the sole focus was on whether a measure proposed by a student was or was not correct. Other challenges included the nature of the conceptual tools available to teachers. Initially, we represented constructs describing the progression of student thinking as tables. These brief descriptions had the virtue of economy but they did not communicate well. Similarly, our initial attempts at lessons were not sufficiently educative – they did not reveal why particular tasks and tools were likely to support student learning. And at first the observation tools were in embryonic form. However, teachers already had a history of exploring the growth of student thinking in other realms of mathematics, especially whole-number arithmetic. As a consequence, our efforts to develop a community of practice centered around student thinking was well received. In this light, we engaged in several forms of community building.

Learning labs

We adapted 'math labs' (Kazemi et al., 2018) to collaboratively generate opportunities to learn from and with students. During a learning lab, teachers collaborated to plan, conduct and reflect upon student learning in situ. Teachers were sometimes grouped by grade band (e.g. K–2, 3–5) and at other times constituted across grades (K–5). An instructional facilitator and I assisted at every lab (two or three labs per day were conducted at each of my monthly visits). The initial phase of the lab included decisions about a portion of a lesson that would serve as a focus. The group anticipated how students might think about this portion – in the language of a construct – and what we were especially interested in seeing in more depth. Occasionally teachers reviewed the mathematical concepts beforehand so that they would be better positioned to interact with students. Usually a pair of teachers conducted the instruction with a class of students while colleagues observed and interacted with small groups of students to characterise student thinking according to sub-levels of one or more constructs. Teachers used the TOTs system to record evidence of student thinking. During the classroom lesson, participants could interrupt or 'pause' activity as needed to draw attention to an unexpected development in student thinking or to propose an alteration in the plan of instruction. During the debriefing sessions that followed, teachers characterised examples of student thinking with respect to the constructs, often displaying samples of student work or replaying instances of student learning. Constructs became tools for dialogue as teachers developed their implications for current and future instruction. Teachers often concluded with plans for future instruction ('next steps'), and/or for modifications to instruction to be enacted in the near future with other classes at the same or other grade levels.

Mathematical investigations

A second form of community building involved group inquiries about the mathematics of measure. For example, teachers investigated properties of dynamic measures of space, such as how a length can be viewed as motion along a path, area as generated by a length moved through a second length, volume as generated by an area moving through a length, and an angle as a directed rotation. They also considered how to help make fractions such as $\frac{7}{3}$ more intelligible to students, and how measurement can be employed to interpret arithmetic operations with fractions, especially multiplication and addition. These investigations were most often conducted in response to teacher requests during summer institutes, but were also a component of many of the learning labs.

Auditing evidence and communal looks at student learning

At the end of the school day during monthly meetings, we jointly examined evidence of student learning that was being generated by teachers, with an eye toward establishing a trail of evidence so that others could access the basis of evidence for a particular assignment of a student to a construct. We compared this process to auditing a tax return. We also used TOTs to consider progress in student learning at grade levels and across grade levels, so that we could visualise school-wide patterns of development. These visualisations instigated conversations about the aspects of instruction that needed further attention. In addition, during these conversations teachers recommended changes to conceptual tools and the TOTs.

Revisions of conceptual tools and TOTs

We engaged in iterative refinement of lessons by adding 'teacher notes' that clarified the instructional intent of tasks and served as guidelines to productive ways of supporting student learning. As noted previously, these were informed by our work together in learning labs. Similarly, as teachers conducted formative assessments, we relied upon the responses to generate guides that a teacher could use to lead productive classroom conversations based on student responses. These guides were subsequently included in lessons. Visualisations and related capacities of TOTs were expanded as teachers used the tool and conversed about progress in student learning during after-school meetings. For example, we added the heat map (Figure 5) and a history function to TOTs to enable teachers to visualise change during the year at multiple grain sizes (class, grade, school). Constructs were revised to include narratives of development, so that teachers could more readily interpret the progress mapped in the tables.

Evidence of teacher and student learning

There are multiple sources of evidence for the robustness of this innovation at different levels of organisation, ranging from district/school to individual teacher and student.

School level

At an organisational level, the innovation is now part of the school's yearly improvement plan and is endorsed by the district as a resource for K–6 mathematics instruction. The building principal has changed, but administrative support for this innovation remains solid. Teacher participation has remained steady with a few additional participants joining during the course of the project. Teacher corner contributions continue to grow, and teachers have insisted on maintaining the learning lab and mathematical investigations components of the community-building enterprise. Statewide summative assessments now suggest that the school is achieving 'value added' in mathematics, especially for those portions of the assessment indicating measurement and rational number.

Teacher level

To gauge growth in a shared professional vision about teaching and learning measure, we conducted flexible interviews on a yearly schedule to inquire about what teachers notice as they observe videotaped lessons about measure, and about their interpretations of the different forms of activity in which they are engaged. We also examine records of learning labs, mathematical investigations, and formative assessment conversations for evidence of growth of professional vision. As an example, we briefly describe change after one year of participation in the professional learning community in what teachers noticed about instruction in measurement.

At the outset of our collaboration with teachers in Sleeve Elementary, teachers viewed three episodes of classroom teaching in measurement. The teaching episodes were drawn from Grade 1, Grade 3, and Grade 5 and were conducted by teachers from a previous research project that investigated longitudinal change in student thinking about measurement. We asked teachers to tell us what they noticed (Sherin et al., 2011) about concepts of measure and about instructional practices with the aim of exploring the growth of professional vision. We solicited teacher noticings again at the end of the first year of our collaboration.

On both occasions we transcribed video and identified segments during which teachers noticed a core concept of measure and/or an instructional practice aimed at fostering student learning. Three overarching classes of codes were employed to characterise what teachers noticed. The first, *Measurement Concepts*, characterised which concepts of measure that teachers tended to notice, such as the need to define an attribute in one episode and the use of dissection to find area measure in a second episode. The second class of code, *Domain-Independent Practices,* described teacher noticings of instructional practices that supported student learning generally by fostering a positive classroom climate. For instance, a participant might mention that the instructor in the video episode encouraged students to share solution strategies or to ask questions. However, these practices were not explicitly related to learning any concept of measure. In contrast, the third class, *Concept-Specific Teaching Practices*, were forms of instructor practice described as helping students learn specifically about one or more of the core concepts in measure. For example, a participant might notice that the instructor employed a metaphor of motion (e.g. travelling a distance, sweeping a length through another length) to help students differentiate between area and perimeter.

We focused on significant transitions between the first and second interviews, which were given one year apart, in what teachers noticed about core concepts and instructional practices. We counted every instance of teacher noticing about instructional practices across all three of the video episodes. At the outset of the project (first interview), teachers most often noticed domain-general practices, which accounted for 54 per cent of noticings about instructional practices. These included instructors' questions ('they are using questioning, and the questions I see were … those higher-level questioning techniques'), instructors' support for student agency ('encourage other students to build upon the thinking of another child'), and instructors' use of materials to support student learning (e.g. 'They are using a lot of visuals'). In contrast, at the second interview, domain-independent noticings decreased to 13 per cent of the total noticings of instructional practice. But noticings of concept-specific instructional practices increased by 61 per cent. And noticings of core measurement concepts increased by 28 per cent, suggesting that teachers were becoming more attuned to coordinating instructional support with identified domain-specific conceptual goals. Table 1 illustrates change in teachers' interpretive framework across all three of the video episodes that they viewed.

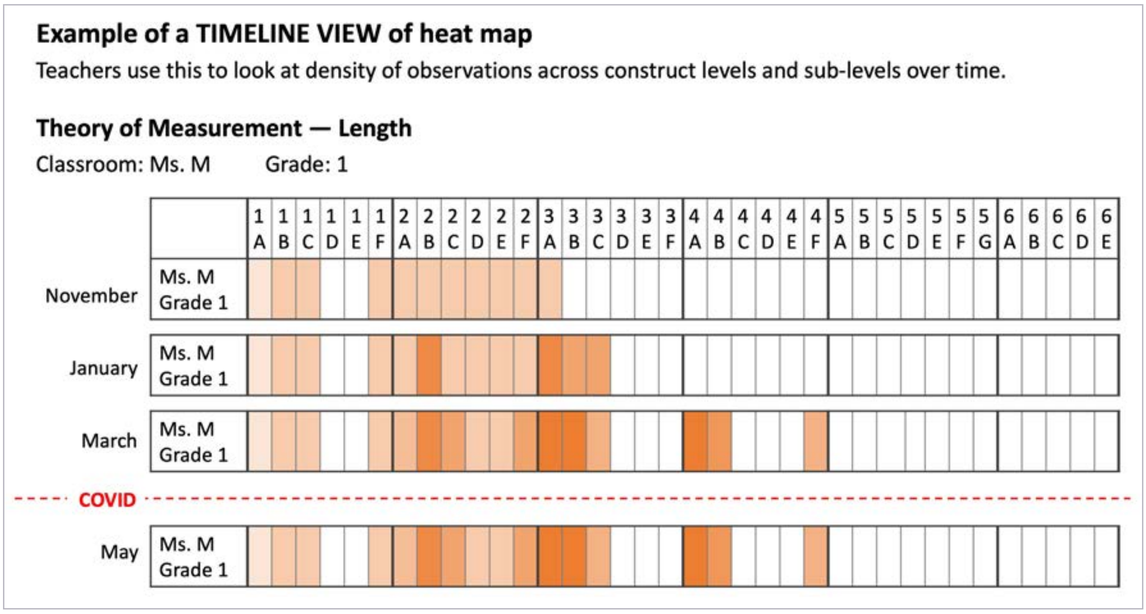
Table 1 Transitions in teachers' interpretive frameworks

Concept/practice noticed		At the onset	One year later
Episode 1 Directly comparing heights and girths of pumpkins Grade 1 students compare the lengths of paper strips generated by different small groups to represent the height of the same pumpkin.			
Concept:	Define attribute	91%	100%
	Direct comparison	18%	82%
	Origin of measure	9%	64%
Practice:	Highlight variability	45%	91%
	Problematise comparison	9%	100%
Episode 2 Finding area and perimeter of an irregular polygon Grade 3 students considered how to find the area and perimeter of a C-shaped polygon figure.			
Concept:	Unit	91%	73%
	Properties of a rectangle	64%	73%
	Dynamic generation of length and/or area	36%	55%
	Differentiation between area and perimeter	36%	55%
	Dissection of area	18%	82%
Practice:	Highlight defining properties of a rectangle	45%	55%
	Appeal to dynamic motion	27%	82%
	Annotate figure	9%	82%
	Gestures to support learning	36%	82%
Episode 3 Interpreting the meaning of a formula for volume measure Grade 5 students interpret the meaning of a familiar formula for the measure of the volume of a prism: length x width x height.			
Practice:	Appeal to dynamic motion	55%	91%
	Tangible model supports visualisation of unit, composite unit (layers)	55%	100%
	Elicit student drawings	18%	91%
	Highlight unit	27%	55%
	Problematise comparison	36%	100%

Student level

The evidence for student learning includes students' responses to summative, construct-based assessments at the beginning and end of every school year. In addition, the predominant form of evidence consists of evidence generated in classrooms of construct-centered growth in conceptions of measure in length, area, volume and angle. This growth is evident in timeline views of heat maps within classrooms and grade level during the year. For example, a timeline view of Mr. M's first grade class during the second year (2019–2020), displayed in Figure 7, can be interpreted as initial understandings by students of the role of measurement in comparing attributes and properties of units, such as tiling (November). The next snapshot indicates an important conceptual transition to understanding unit iteration (3A) and symbolisation of units (e.g. 0, 1) on a scale, (3B, C), by mid-year, and then further progress toward part-unit iteration (4A) and location of part-units on a scale (4B) by early spring (March in the Northern Hemisphere). Further evidence of learning was interrupted by school closure due to the COVID-19 pandemic.

Figure 7 History of learning about length measure in a Grade 1 class



Discussion

Fostering practices of assessment so that they serve as routine guides to teaching and learning is a goal of most programs of ambitious instruction in mathematics. Knowledge of student thinking and of typical horizons of change are repeatedly cited as critical components of teacher knowledge that undergird adaptive instruction (e.g. Copur-Gencturk et al., 2019; Gibbons & Cobb, 2017). Yet even though teachers' ongoing assessments of student learning are vital to instruction, they are not routinely incorporated into systems of assessment that are used for accountability purposes. To do so, we have identified a set of resources that we believe are vital for bringing teacher voice to larger-scale, summative assessment. One resource is organisational – the need to institutionally support continued teacher learning and collaboration. In this project, we have adapted the math lab approach to continuous improvement of teaching and learning so that assessment practices become strongly coupled to student (and teacher) learning. Instruction is informed by continuous formative assessment, with an expanded sense of what constitutes an 'item' in the traditional sense of assessment. Of course, this kind of continuous assessment would not be possible without tools like TOTs, which afford capture of student thinking and visualisation of progress at multiple levels of inquiry.

A second resource consists of a common language of learning that can be employed to interpret student responses in a variety of settings. In this project, these are manifested as constructs, which are representations cast at an intermediate level of description. The level of description is chosen to be noticeable as professional vision (Goodwin, 2018) develops, and to be actionable, in the sense that the construct description of student thinking is specific enough to warrant instructional support. Instructional support is assisted by curricular tasks and tools, especially as these are deployed during learning labs. The ensemble of curricular co-design, routine practice of formative assessment embedded in ongoing classroom activity, and a community of practice support children's and teachers' learning (as well as those of us from the university).

The fact that they are designed with common constructs in mind does not necessarily imply that student performances on summative tests and in classroom tasks will be identical. We do not conceive of students as having or not having a particular property that is being measured, but instead think of students as manifesting particular understandings in particular settings. That is, measurement of qualities of thinking is entangled with the circumstances of its generation. What we anticipate is that with constructs, we can interpret student responses consistently across settings and tasks, taking into account variation in circumstances of performance. We are still in the midst of this innovation, so more definitive relations between summative and classroom assessment are still being investigated.

The research reported was conducted in collaboration with Leona Schauble, Corey Brady, and Panchompoo Wisittanawat, and with our K–5 teacher partners at Sleeve Elementary and Mallard Elementary. Funding was provided by the National Science Foundation, DRK-12 program, # 1621088.

References

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In L. Darling-Hammond & G. Sykes (Eds.), *Teaching as a learning profession: Handbook of policy and practice*. (pp. 3–32). Jossey-Bass.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, & Practice*, *5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation, and Accountability*, *21*, 5–31. https://doi.org/10.1007/s11092-008-9068-5

Copur-Gencturk, Y., Plowman, D., & Bai, H. (2019). Mathematics teachers' learning: Identifying key learning opportunities linked to teachers' knowledge growth. *American Educational Research Association*, *56*(5), 1590–1628. https://doi.org/10.3102/0002831218820033

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. https://doi.org/10.3102/0013189X08331140

Gibbons, L., & Cobb, P. (2017). Focusing on teacher learning opportunities to identify potentially productive coaching activities. *Journal of Teacher Education*, *68*(4), 411–425. https://doi.org/10.1177/0022487117702579

Goodwin, C. (2018). *Co-operative action*. Cambridge University Press.

Horn, I. S., Kane, B. D., & Wilson, J. (2015). Making sense of student performance data: Data use logics and mathematics teachers' learning opportunities. *American Educational Research Journal*, *52*(2), 208–242. https://doi.org/10.3102/0002831215573773

Kazemi, E., Gibbons, L. K., Lewis, R., Fox, A., Hintz, A. B., Kelley-Petersen, M., & Balf, R. (2018). Math labs: Teachers, teacher educators, and school leaders learning together with and from their own students. *NCSM Journal of Mathematics Education Leadership*, *19*(1), 23–36.

Kim, M. J., & Lehrer, R. (2015). Using learning progressions to design instructional trajectories. In C. Suurtamm & A. R. McDuffie (Eds.), *Assessment to enhance teaching and learning* (pp. 27–38). The National Council of Teachers of Mathematics.

Lehrer, R., & Schauble, L. (2015). Learning progressions: The whole world is not a stage. *Science Education*, *99*(3), 432–437. https://doi.org/10.1002/sce.21168

Wiliam, D. (2015). Assessment: A powerful focus for the improvement of mathematics instruction. In C. Suurtamm & A. R. McDuffie (Eds.), *Assessment to enhance teaching and learning* (pp. 247–253). The National Council of Teachers of Mathematics.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.

Rethinking measurement for accountable assessment

Professor Mark Wilson
University of California, Berkeley & University of Melbourne

https://doi.org/10.37517/978-1-74286-638-3_13

Mark Wilson is a professor of education at the University of California, Berkeley, and at the University of Melbourne. He teaches courses on measurement in the social sciences, especially as applied to assessment in education. He was elected President of the Psychometric Society, and, more recently, President of the National Council for Measurement in Education (NCME). His research and development interests focus on the development and application of sound approaches for measurement in education and the social sciences, the development of statistical models suitable for measurement contexts, the creation of instruments to measure new constructs, and scholarship on the philosophy of measurement.

Abstract

The underlying model for most formal educational measurement (e.g. standardised tests) is based on a very simple model: the student takes a test (possibly alongside other students). The complications of there being an instructional plan, actual instruction, interpretation of the outcome, and formulation of next steps, are all bypassed in considering how to model the process of measurement. There are some standard exceptions, of course: a pre-test/post-test context will involve two measurements, and attention to gain score, or similar. However, if we wish to design measurement to hold to Lehrer's (2021) definition of 'accountable assessment' – as 'actionable information for improving classroom instruction' – then this narrow conceptualisation must be extended. In this presentation, I will posit a simple model that reflects the simple one-test context described above, and then elaborate on it by adding in a) a framework for design of the assessments that is keyed to educational interpretation, b) further rounds of data collection that can indicate changes in a student's underlying ability, and c) provision for varied assessment modes that will allow for i) classroom-independent tasks that operate at the summative and meso levels, and ii) classroom-dependent tasks that operate at the micro level. The former are designed to provide a basis for triangulating student responses across different contexts, and the latter are designed to closely track the variation of student performance over time in a classroom instructional context. This framing will be exemplified in a in a K–5 elementary school that is seeking to improve the quality of instruction and students' understandings of measure and arithmetic. The different levels of data collection will be instantiated by two different pieces of software, which operate at the micro level and the meso/summative levels respectively.

Introduction

The underlying educational context for educational measurement has generally been one where a student is seen as progressing through an instructional plan delivered by their teacher(s); the student (along with their peers) is then subjected to a test designed to assess the expected range of outcomes, and the teacher(s) then plan for the next instructional step based on that assessment.

Of course, there may be further rounds of this – retesting, etc. Although this rather general formulation is well-known to most involved in measurement, the actual paradigmatic context on which educational measurement is predicated is much simpler – the student takes a test (possibly alongside other students). The complications of there being an instructional plan, actual instruction, interpretation of the outcome, and formulation of next steps, are all bypassed in considering how to model the process of measurement. There are some standard exceptions, of course – a pre-test/post-test context will involve two measurements, and attention to gain score, or similar.

We see these types of testing as being representative of the macro level of assessment. The *macro level*, commonly also called 'summative testing', is the level of most standardised tests. The topics are relatively coarse composite constructs, such as science, language arts, geometry, and the tests are largely used for administrative decisions by parents and administrators/policymakers. They are used over relatively longer education time-periods (years, semesters, program length, etc.) for relatively large-scale decision-making, such as passing a course, grade advancement and course-placement.

However, if we wish to design measurement to hold to Lehrer's (2021) definition of *accountable assessment* as 'actionable information for improving classroom instruction', then this narrow conceptualisation must be extended. In this presentation, I will posit a model that starts with the simple one-test context described above, and then elaborate it by adding in 1) a framework for design of the assessments that is keyed to educational interpretation, 2) further rounds of data collection that can indicate changes in a student's underlying ability, and 3) provision for varied assessment modes that will allow for a) classroom-independent tasks that operate at the macro (summative) and meso levels, and b) classroom-dependent tasks that operate at the micro level. The *meso level* assessments are designed to provide a basis for triangulating student responses across different contexts, and the *micro level* assessments are designed to track closely the variation of student performance over time in a classroom instructional context.

This framing is exemplified in work from a K–5 elementary school that is seeking to improve the quality of instruction and students' understandings of measure and arithmetic. The different levels of data collection will be instantiated by two different pieces of software, which operate at the micro level and the meso/macro levels respectively.

The National Science Foundation Collaborative Research Project *Modeling assessment to enhance teaching and learning* (MAETL) is a collaboration among Richard Lehrer, Leona Schauble and Corey Brady from Vanderbilt University, and Mark Wilson and Perman Gochyyev from the University of California, Berkeley. The purpose is to create and test-out a novel assessment system designed to address two coordinated purposes:

1. to provide ongoing, instructionally productive evidence to teachers about student learning in the context of learning progressions
2. to link dense information from student work products and formative assessments (meso and micro assessments) in new models that generate robust estimates of the growth of student learning.

The specific topics of instruction are Measurement of Length, Area, Volume, Angle, and Measured Quantities (as entrée to Rational Numbers – Fractions as quantities, fractions as operators). Consider the first construct, Theory of Measurement – Length (ToML) as an example; this construct describes how children come to constitute a theory of measure to compare magnitudes (extents) of lengths. and is represented using a 'construct map', based on the Bear Assessment System (BAS; Wilson, 2005) and is illustrated in Figure 1. The levels of a construct map are designed to encapsulate important qualitative steps towards the highest level.

For example, at the second level, ToML 2, students focus on the nature of a unit. They must learn (in practice) that:

1. units enable indirect measurements via accumulation and count (instead of directly comparing), and
2. units allow for both additive (how much longer?) and multiplicative comparisons (how many times longer?).

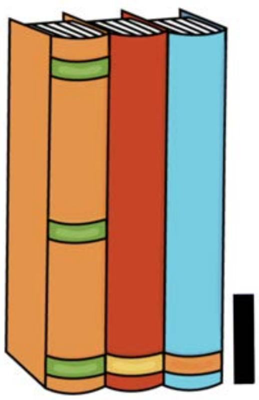
Figure 1 The construct map for TOML

6. Generalising relationships (e.g. Measure of A in B is the reciprocal of measure of B in A)
5. Partitioning and symbolising involving 3-splits and composition of 2- and 3-splits
4. Partitioning, iterating, symbolising partitioned units – 2-splits
3. Iterating units and symbolising distance travelled
2. Explaining properties of units and their role in accumulation
1. Directly comparing

Students must develop understandings of the properties that enable these uses. Hence, students at this level need to be able to explain the roles of identical units and tiling. Then, as they move up to ToML 3, they must show that they can use these units to measure something, such as is exemplified in the item shown in Figure 2. The items are developed and delivered using the BEAR Assessment System Software (BASS; Wilson et al., 2019).

Figure 2 An item at level 3: Height of a Book

Carlos started measuring the height of the blue book, but he does not have enough units. Carlos says he cannot finish measuring the height.

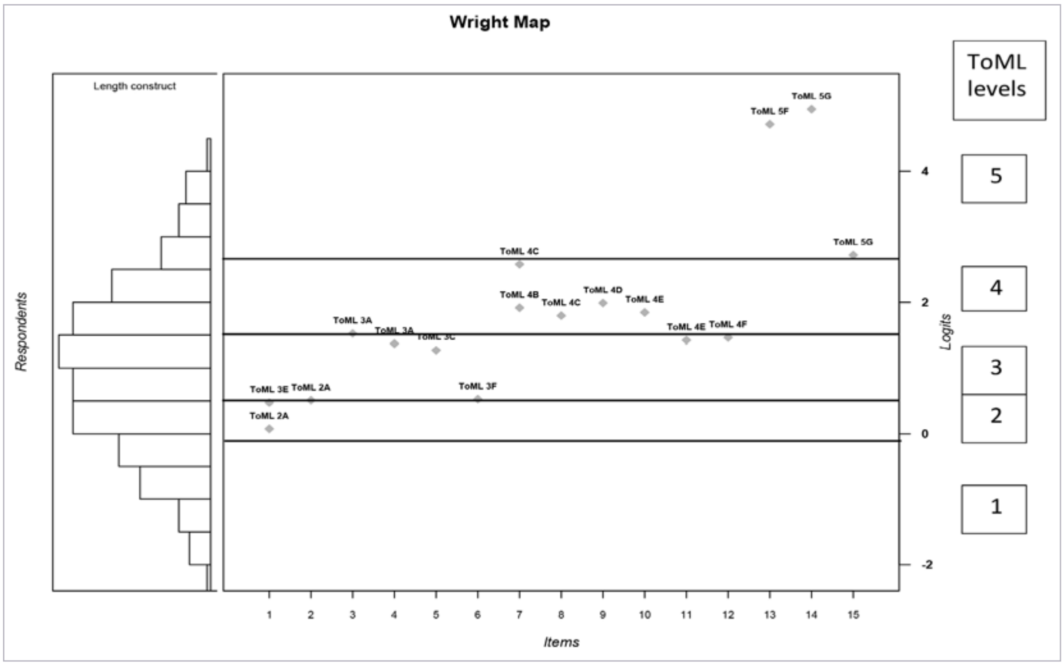


Do you agree with Carlos?

- Yes
- No

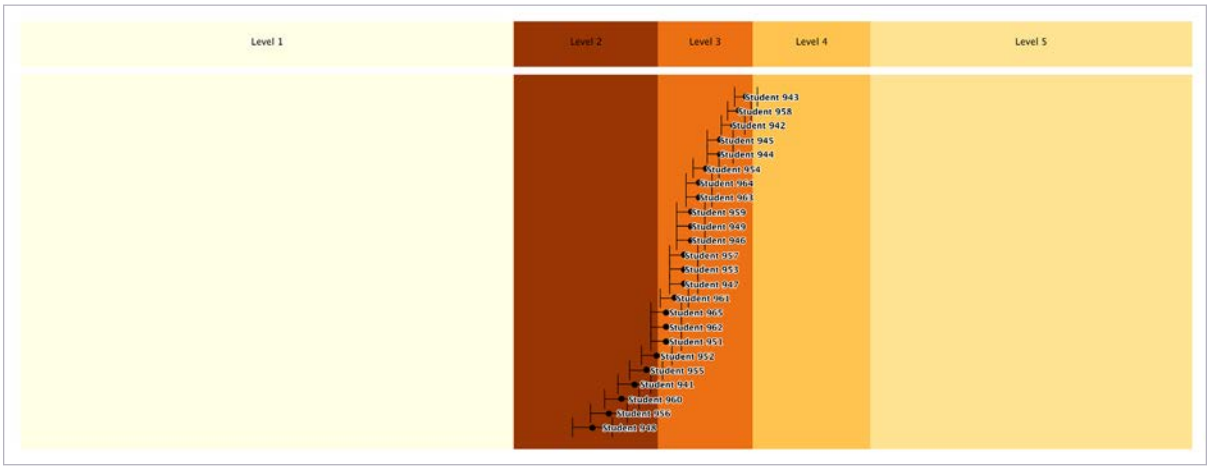
Items such as these are designed to be used at the meso level. This is the level of testing for teaching and tends to deal with the contents of broader standards. Specifically, it relates to what one might call 'teachable constructs', such as, buoyancy, variability or measurement. They can be used by teachers to put together tests, and the results would be used by teachers, and, when old enough, students themselves. Typically, they might be used at instructional-scale time-periods and up (i.e. days, weeks, teaching units, etc.) for making instructional decisions such as planning for a day/week's topics, what topics to revisit, which students need extensive help. Results for the ToML items such as these can be displayed by BASS as Wright Maps, and the locations of the item thresholds can be graphically summarised. A Wright Map is a graph that simultaneously shows estimates for both the students and items on the same (logit) scale, an example of which is shown in Figure 3. On the left side of the Wright Map, the distribution of student abilities is displayed as a histogram (on its side), where ability entails knowledge of the skills and practices for ToML. The person's abilities have a roughly symmetric distribution. On the right side are shown the thresholds for 15 items in ToML (where two of the items are polytomous – that is, the student responses can indicate more than one level of the ToML construct map). Each item is represented as 'ToML Lk', where L represents the ToML Level of that item, and k represents a finer-grade coding not discussed in this brief paper. Looking beyond the results for a single item, we note the consistency of the locations of these thresholds across items. We used a standard-setting procedure called 'construct mapping' (Draney & Wilson, 2011) to develop cut-scores between the levels. Following that process, we found that the thresholds fall quite consistently into the ordered levels from ToML 1 to ToML 5, with a few exceptions, at the borderlines between the levels.

Figure 3 Wright Map for Theory of Measurement – Length



Having established these cut points between ToMl Levels, we also can use purpose-built BASS reports, such as the Group Report shown in Figure 4, which lay out the distribution for a whole class across the ToML Levels 1 to 3. In these reports, the estimated student location is noted as a black dot, and a 66.7 per cent (i.e. 1-sigma) confidence region is shown around that location. Here we can see that this class extends from Level 2 to Level 3, just bordering on Level 4.

Figure 4 Group report for a (fictional) class



The construct map can be used to tie in the meso level assessments with the micro level in the classroom. At the micro level, relatively informal and fine-grained assessments are used for within-instruction observations. These typically relate to relatively fine-grain sized knowledge, 'in-pieces' such as the definition of density, initial experiences of variation, what it means to measure a length, etc. They are associated with the opportunities teachers have for telling observations as they walk around their classrooms and are focused on brief education time-periods (i.e. a sequence of concepts within a classroom unit, etc.), and are intended to inform smaller-scale instructional decisions, such as what tactic to use next in discussing a certain idea with students. If one were to be developing computerised teaching software, then this is essentially the same level as the software would be operating at.

The levels of the construct map are the key to bringing these two levels of information together. The project has developed Teacher Observation Tools (TOTs), which is a mobile data-gathering iPad application designed for teacher use while teaching in their classrooms. A sample screenshot is shown in Figure 5. To use this software, a teacher must learn the construct map (in this case the ToML construct map) well enough to apply it 'on the fly' in their classroom – note the selection of ToML3A for the recorded classroom event. An example of the data records for one classroom is shown in Figure 6. These data records can be input to the BASS database and a sample group report is illustrated in Figure 7.

Figure 5 A screen shot from TOTs

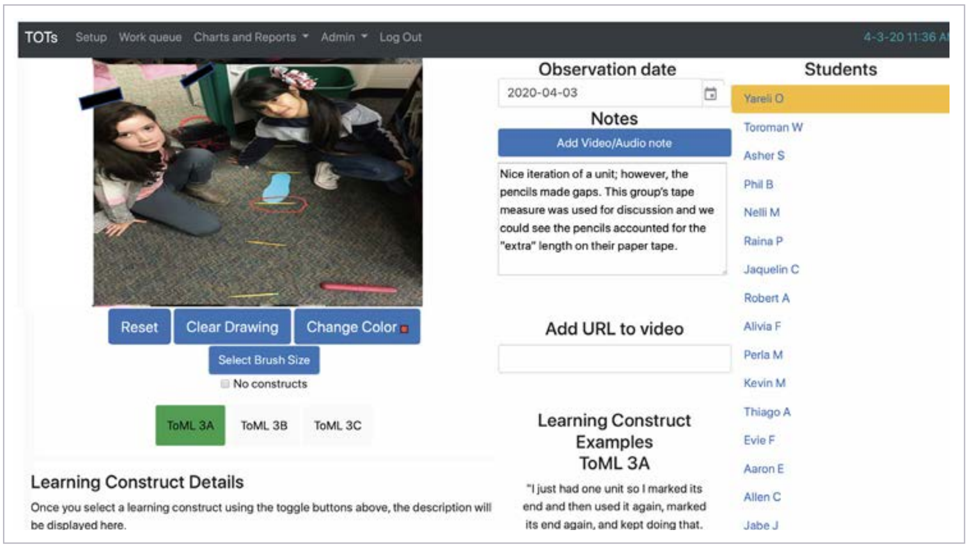


Figure 6 TOTs record for across time for one class

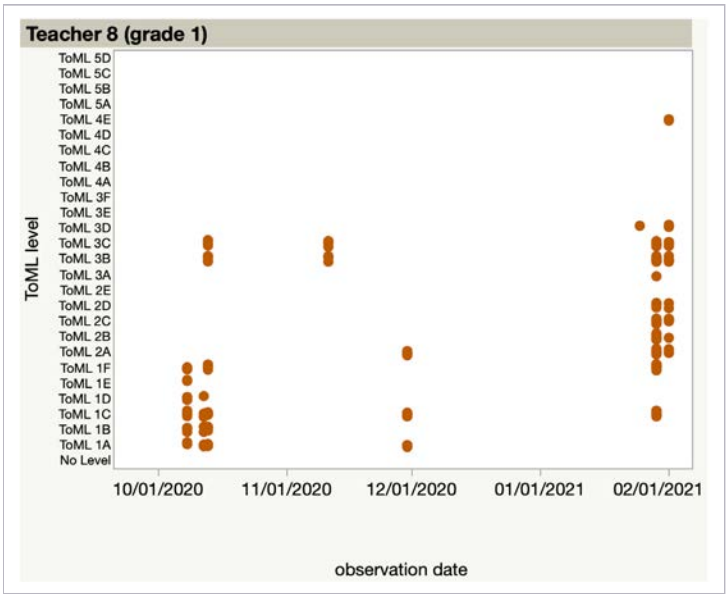
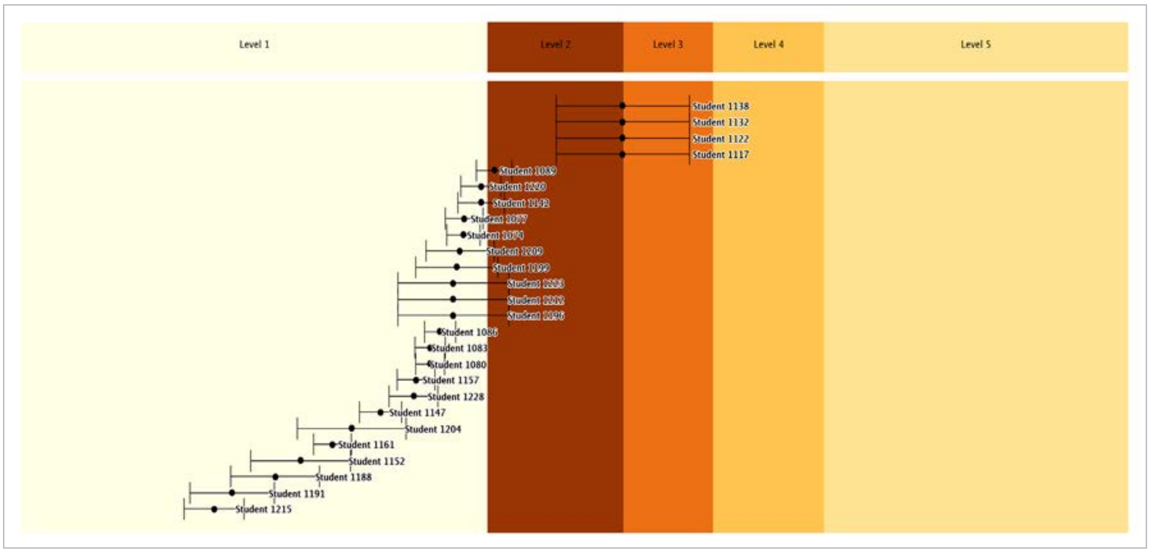
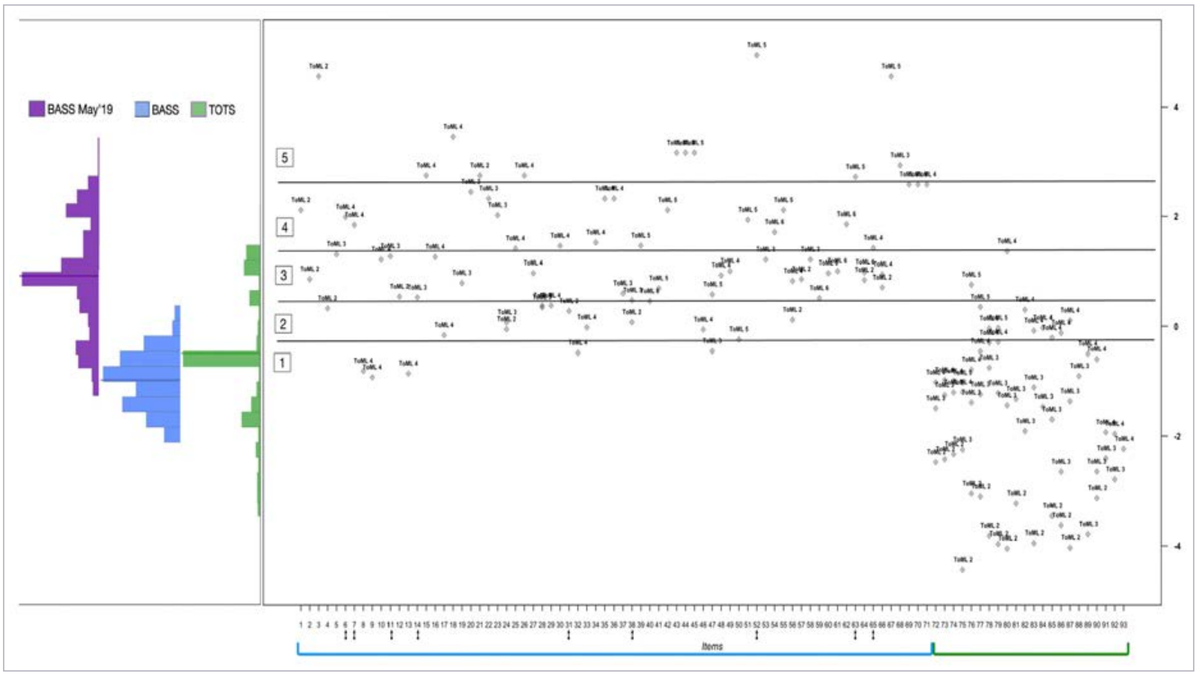


Figure 7 BASS group report for the TOTs data



The two levels are coordinated via the levels of the ToML construct map, and representative results from scaling the two together are shown in the expanded Wright Map in Figure 8, where the BASS levels illustrated in Figure 3 are used to interpret the findings. These ToML Levels were established in the 2019 post-test, for which the student estimates are shown on the left-hand side in the purple histogram. As we can see, the students, who ranged from Grade 1 to Grade 5, spread along the scale from ToML Level 1 to 5. Next to the right is a blue histogram showing the student estimates at the pretest in 2020 – as one might expect, these are much lower than for the previous year's post-test results, ranging up to only Level 2. The next histogram (green) shows the TOTs estimates for students in the first part of the 2020/21 academic year (i.e. the same students as shown in the blue histogram), and here we see an interestingly broader range than for the pretest. This increase in breadth can be attributed partly to 1) the initial instruction in the program, but also to 2) the scaffoldings to student performance provided in the classroom, and 3) the increased appreciation of teachers for the communications of their own students. One interesting extra feature is the locations of the TOTs items on the right-hand side of this Wright Map: first, it is shown as a 'cloud' of micro-level items, as we modelled them as random-effects, and second, the locations are all lower than for the meso-level BASS items, and we see this as being attributable to, again, 1) the scaffoldings to student performance provided in the classroom, and 2) the increased appreciation of teachers for the communications of their own students.

Figure 8 Reconciling the meso and micro levels of results



In conclusion, we note that the conceptualisation of 'accountable assessment' involves the matching development of assessment at both the meso and micro levels of assessment, as exemplified in this brief paper. The BASS and TOTs software accommodate these assessment levels, and the possibility of coordinating between the two is accomplished by basing *both* on the relevant construct map (in this case ToML). Other approaches to this conceptualisation are also possible (e.g. Doignon & Falmagne, 1999), though generally they are built only at the finest (micro) grain size, which, while needed for applications such as computer-based teaching, may not serve human teachers so well.

References

Doignon, J. P. Falmagne, J. C. (1999). *Knowledge Spaces*. Springer-Verlag.

Draney, K., & Wilson, M. (2011). Understanding Rasch measurement: Selecting cut scores with a composite of item types: The Construct Mapping procedure. *Journal of Applied Measurement*, *12*(3), 298–309.

Lehrer, R. (2021, August 16–20). *Accountable assessment*. [Keynote presentation]. Australian Council for Educational Research Conference (online).

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Erlbaum.

Wilson, M, Scalise, K., and Gochyyev, P. (2019). Domain modelling for advanced learning environments: the BEAR Assessment System Software. *Educational Psychology*, *39*(10), 1199–1217. https://doi.org/10.1080/01443410.2018.1481934.

Developing an assessment of oral language and literacy: Measuring growth in the early years

Dr Dan Cloney
Australian Council for Educational Research

Kellie Picker
Australian Council for Educational Research

https://doi.org/10.37517/978-1-74286-638-3_2

Dr Dan Cloney is a Senior Research Fellow in the Education Policy and Practice Program and a Member of the Global Education Monitoring (GEM) Centre at the Australian Council for Educational Research (ACER). Dan's expertise is in early education, cognitive development, and academic achievement.Dan's research program focuses on the potential for high-quality early childhood education and care programs to support all children to flourish and to reduce development gaps caused by inequality. Current studies include the Overcoming Disadvantage in Early Childhood study (in partnership with the Australian Literacy & Numeracy Foundation) evaluating the effectiveness of the Early Language & Literacy program in New South Wales, and the Modelling of Universal Pre-Primary Education Study (in partnership with UNICEF) evaluating the effectiveness of pre-primary programs in Bogor, Indonesia. He is also an Honorary Research Associate at the Murdoch Children's Research Institute (MCRI) and the Melbourne Graduate School of Education (MGSE).

Kellie Picker is a researcher for the Australian Council for Educational Research (ACER). She is currently working on the Overcoming Disadvantage in Early Childhood study evaluating the effectiveness of the Early Language & Literacy program in New South Wales, as well as the Learning Through Play in School research project. The Learning Through Play at School project is an intervention study aiming to help Ukrainian teachers transition from adult-centred teaching to child-centred learning through the use of play pedagogies. Kellie's expertise is in early childhood education, early literacy development and effective pedagogies. She was a researcher on the Western Australian Teaching for Growth *project and has been a classroom teacher, a learning enrichment specialist and an IT coordinator in primary education. Kellie is in the final stages of completing a PhD at the University of Melbourne with a focus on deepening our understanding of the role of teacher reading content knowledge in early primary education.*

Abstract

Children develop rapidly in their early years. A crucial component of this development is a child's ability to learn and use language. Even before they enter formal education, children have learned much about oral language and literacy through meaningful interactions with others, and from their life experiences. Children, however, do not develop at the same pace – some children arrive in early childhood education and care (ECEC) programs more advanced while others require additional support. Recent reviews of the assessment tools available to ECEC educators show a lack of good quality measurement and a reliance on checklist style inventories or narrative approaches. This paper presents a new measure of oral language and pre-literacy specifically designed to be accurate enough to reliably measure an individual child's growth. Results from a combined calibration of children's responses using a many-facets item response model show the measure to be reliable, valid and sensitive enough to measure growth within children and between groups of children over time. Implications for future assessment development and for educators' practice are discussed, including how such measures can provide insight into what children know, understand, and can do (Reynolds, 2020) and what educators can do to support future learning experiences targeted at children's specific language and literacy needs.

Introduction

Oral language and literacy

The development of oral language and literacy is a core skill, important in its own right and also predictive of later learning and development in both school curriculum areas (academic achievement) and in general capabilities, such as cognitive and social skills. Oral language research highlights that it is our innate need and ability to communicate using language that supports and develops our use of expressive and receptive language. Here, research suggests that children who start formal education with advanced oral language skills, are more likely to be successful readers (Foorman et al., 2015).

Literacy learning, including oral language, does not come naturally, and children require specific knowledge and skills so they can learn to read and write. Underpinning the construct of literacy are five key elements:

- phonemic awareness (a sub-strand of phonological awareness)
- phonics
- fluency
- vocabulary
- comprehension.

These five elements were identified by the National Institute of Child Health and Development (NICHD) (2000) and Rowe (2005) as essential to the effective teaching of reading in English speaking classrooms and have come to be known as the 'science of reading'. However, in a review of the literature, Konza (2014) argued that due to oral language being foundational to literacy learning, it should be included as an essential element of the science of reading, renaming them the 'Big Six'.

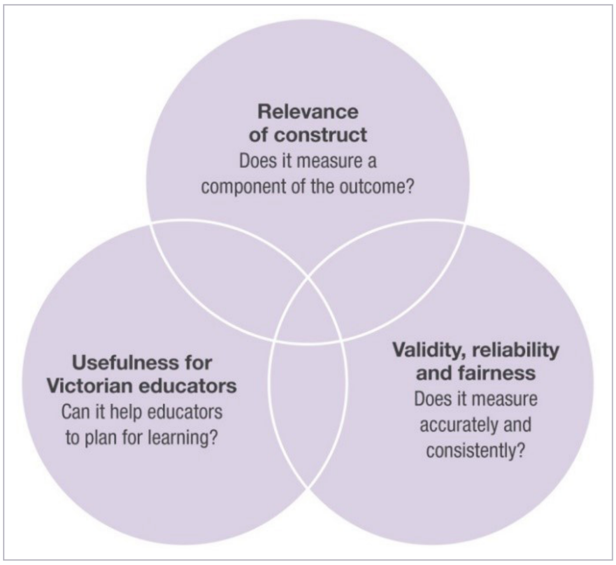
In Australia, the 'Big Six' are represented in the national learning progressions. Here, 'oral language, vocabulary and comprehension are reflected across many areas of the progressions, and phonemic awareness, phonics and fluency are addressed as sub-elements' (Australian Institute for Teaching and School Leadership [AITSL], 2020, p. 8).

Assessment

There is a lack, however, of quality assessments that measure language and literacy and describe how they develop over time. Recent literature reviews of assessment available to early educators show that assessment tends towards checklist-style inventories that are designed to screen for developmental problems and do not describe what children can do and what they might do next (and what educators might do to support this). While other forms of assessment tend toward narrative that is difficult to communicate to other educators, parents and to teachers at transition, and is applied inconsistently (Anzai et al., 2021; Cloney et al., 2020).

Assessing children's language and literacy growth is complex and multifaceted, particularly in early childhood where assessment cannot take on the form of traditional school age assessment (completing a test) but rather should be embedded within authentic interactions between adults and children. Cloney et al. (2020) developed a series of principles that help educators understand the elements of quality assessments in the early years, as shown in Figure 1. These principles are anchored in best practice for early learning, informed by frameworks such as the Early Years Learning Framework and the Victorian Early Years Learning and Development Framework.

Figure 1 Principles for assessing oral language and early literacy in children, adapted from Anzai et al. (2021, p. 23)



Principles for good quality assessment in the early years

The first two key principles apply to the selection of an assessment, to ensure it addresses the knowledge, skills and concepts (construct) being measured, in a way that is developmentally appropriate for the age and stage of the children. It is important that educators are familiar with how these constructs are discussed in the literature, to assist the educator in making decisions about the suitability of the assessment. This includes the breadth and/or depth of coverage of the construct and whether it aligns with framework and curriculum documents that govern teaching and learning. Such an assessment should also be designed to maintain or even enhance those strong, warm and responsive relationships that are foundational to children's learning (Pianta et al., 2008). Along with maintaining relationships, effective assessments engage children in meaningful interactions and experiences that are challenging but within the child's learning reach (Palermo et al., 2007).

The next principles of assessment consider the validity, reliability and fairness of the measure. Assessments are found to be 'valid' when they accurately measure the construct/s they claim to measure, for example, expressive language. 'Reliable' assessments are said to produce 'valid results consistently across contexts' and measures of 'fairness' are based on whether an assessment provides children with unbiased opportunities to express what they know, without disadvantaging specific groups of children (Cloney et al., 2020, p. 19).

Finally, it is of the utmost importance that the information collected in an assessment can be used by educators to understand and describe student's growth on a continuum of learning. Educators who have a deep understanding of how language and literacy develops and can discuss growth as a continuum of learning, such as those described in learning progressions, are well equipped to facilitate children's learning. Educators with knowledge about what children know and can do now, can use assessment information to plan for what children need to know and do next. This in-depth knowledge means educators can target learning experiences, scaffold children's learning, and contribute to the growth of all children (Cloney et al., 2020).

Research questions

This paper therefore asks whether it is possible to develop an assessment that is contextually appropriate for children (built around authentic tasks and one-on-one interactions between an adult and child) and still reliable, valid and fair. That is, can best-practice in assessment development and measurement be applied to an early years measure? Further, is such a measure sensitive enough to measure growth within children and between groups of children?

Method

Participants

Twenty-five early childhood education and care (ECEC) services were sampled from one regional area of NSW. Twelve of these were sampled from ECEC services implementing a specific oral language and literacy intervention (the intervention group) and 10 were matched (on observed characteristics including National Quality Standard (NQS) rating, service type, size, neighbourhood socioeconomic background) controls (the control group).[1] In total, 22 agreed to participate, with one service being uncontactable (likely closed), one having fewer than 10 enrolments, and one refusing. From these services, preschool-aged children (generally children in the age range 3–5 years) were invited to participate. In total, 571 children in 27 rooms or groups within the recruited centres agreed.[2] The final achieved sample is described in Table 1 and Table 2, disambiguated by the intervention and control groups.

Table 1 Service-level characteristics achieved sample

Variable	Intervention	Control
Count	12	10
Average enrolment	42	41
Exceeding NQS	3	5
Meeting NQS	6	4
Working towards NQS	3	1
Count community preschool	8	6
Count long day care	4	4
Median SEIFA IRSAD*	863	923

* Socio-Economic Indexes for Areas, Index of Relative Socio-Economic Advantage and Disadvantage

1 Although not a focus of this paper, it is important to note that the centres are not a random sample of all ECEC services in the region nor is the sample representative of the services operating in the region.

2 All classrooms or groupings of children were recruited to the Study where they included at least five children of preschool age (e.g. who would be eligible to attend school in 2019, whether their parents intended them to or not). Within each sampled room or group, all children and their main caregiver (a caregiver who lives with the child, and usually the person who drops them off to the centre, or the person who knows the most about the child and their daily routine) were invited to participate.

Table 2 Counts of participants in achieved sample

Variable	Intervention	Control
Centres	12	10
Rooms/groups	14	13
Educators	14	13
Children	263	308
Main caregivers	263	307

Instrumentation

The purpose of this study was to generate a measure of oral language and literacy that was suitable to measure within-child growth and that represented a developmental continuum that covered the range of abilities of children from age (approximately) 2 to 8 years. The assessment was designed in line with the Big 6 framework and assessment principles described in the introduction.

Items were developed around a range of interactions, including a picture book reading comprehension and extended response activity (children are read a picture book and asked a range of questions including extended response), a scene picture (a familiar scene with actions happening that the child is asked to observe and or discuss; for example see Figure 2) and more traditional question-and-answer items (a child is given a prompt and asked to respond). Some responses were audio recorded for the purpose of later coding against a rubric to allow deeper understanding of children's oral language – particularly concepts around sophistication of ideas, fluency, and clarity of expression.

Figure 2 Beach scene oral language stimulus, interview script and scoring guide

1	**Instructions and Questions** *Pass the child the beach scene.* Now please hold this picture and look at it carefully. Then I will ask you some questions about it. *Give the child 30 seconds to look at the picture.* There are lots of people doing things. Tell me something that is happening in the picture. *If no response, prompt with:* Can you tell me what someone is doing? *If the child shakes head then stop; if nods head then say:* Could you please tell me, even in a whisper is ok.			
		Responds with an observation about the picture **1**	Responds but unrelated to the picture **0**	Child says I do not know or no response **99**
	Expressiveness	◯	◯	◯

In addition to the main assessment instrument, contextual questionaries and publicly available data were used to elicit demographic and background information about children, families, home environments, ECEC programs, and local communities.

Analytic approach

All analysis was conducted in ACER ConQuest Version 5 (Adams et al., 2020) and the conquestr package (Cloney & Adams, 2021) in R (R Core Team, 2020).

Calibration

With up to two observations of each child in the study, all information is used in a concurrent calibration to yield item parameters. To do this, the data are represented as a single row per child and cycle combination. That is, children who were observed twice have two rows in the data set.

We estimate a one parameter item response model (1PL) – the many facets model (facet model) (Linacre, 1993) – an extension of the Partial Credit Model (PCM) (Masters, 1982). This facet model allows the responses (at all time points) to all the items, to be decomposed into an item difficulty component (the location of the items on the oral language continuum) and some average deviation from that difficulty at each time point. In this model, the assumption is that as time increases, the items get easier.

Child responses to items are integer scored from 0 (most incorrect) to m (most correct) at each time point t. If we denote the latent ability of child n as θ_n, and the 'difficulty' of each item, i, is made up by the item category boundaries (e.g. the boundary between scoring 1 rather than 0, and between scoring 2 rather than 1), which have three components, δ_i (the 'average' difficulty of the item), plus $\tau_{i\kappa}$ (the deviation from the average difficulty for this category boundary), plus α_t (the average change in the difficulty of items over time), then the probability of child n scoring x on item i at time t is given by:

Equation 1 Probability model of the many facets item response model

$$p(X_{nit} = x) = \frac{exp\ \Sigma_{j=0}^{x}(\theta_n - (\delta_i + \tau_{i\kappa} + \alpha_t))}{\Sigma_{\kappa=0}^{m} exp\ (\Sigma_{j=0}^{\kappa}(\theta_n - (\delta_i + \tau_{i\kappa} + \alpha_t)))}$$

The continuing product of the probabilities for child n's responses to many items represents a likelihood, given their response vector. The unknown parameters θ[3], δ_i, $\tau_{i\kappa}$ and α_t are then estimated by maximum likelihood.

Scaling and modelling

Taking the item parameters (δ_i, $\tau_{i\kappa}$) from the calibration stage as fixed, a two-dimensional 1PL item response model (one dimension for each time point) is estimated. That is, the time facet is removed from the specification and each time point is modelled as a dimension. This yields a measurement model where the increasing ability is reflected in the change in the response vectors (i.e. older children tend to get more items correct) holding the item difficulties constant (note that only some items are used between time points). This is equivalent to a latent growth model, as there is some fixed (average) growth between time point 1 and 2, and a random effect within children at each time point (some random deviation from the average at each time point). This model is used to calculate

3 The distribution of θ involves a population distribution assumption (estimation is by marginal maximum likelihood) that means the parameters of the model are influenced by both the responses to the assessment items as well as the regression variables in the model.

factor scores (plausible values [Von Davier et al., 2009]) at each time point that are then used in secondary analysis. A limited set of conditioning variables are included in the model (population model) to support the secondary analysis in this paper – including a variable indicating whether children are in the intervention and control group – a discussion of the generation of plausible values with regression variables is outside the scope of this work (e.g. Wu, 2005).

Results

Calibration

In general, the model shows good fit to the data (Adams & Wu, 2009). Most items fit between rule-of-thumb guides for weighted mean square (WMNSQ). An example item characteristic curve is given in Figure 3 – in this figure the WMNSQ (sometimes called infit) is 1.06 (95%CI 0.87–1.13). A plot of the item (deltas) fits against the quantiles of the normal distribution (a Q-Q plot) reinforce this – the expectation is that WMNSQ is distributed with expectation 1 and variance 1. Note that in Figure 4, 10 items showing poor fit (all underfit) to the model are excluded.

Figure 3 Example item characteristic curve for a polytomous item with good fit to the model

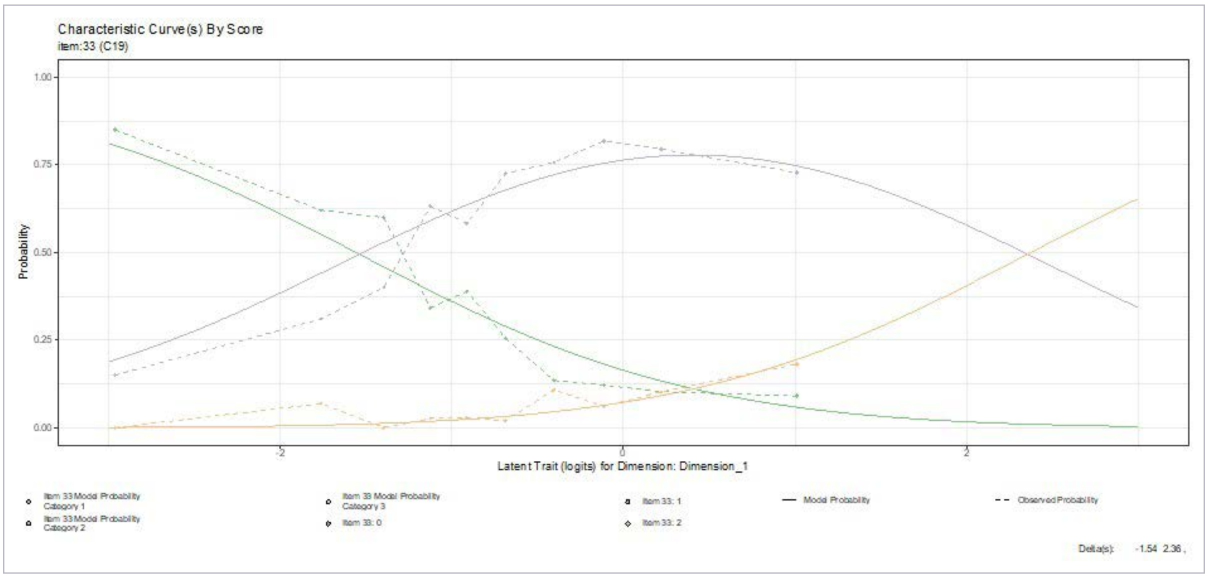
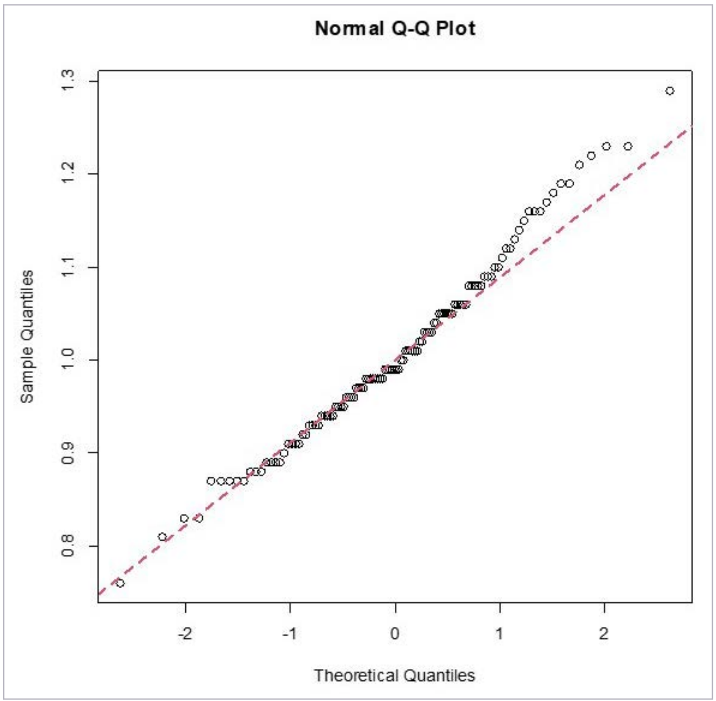


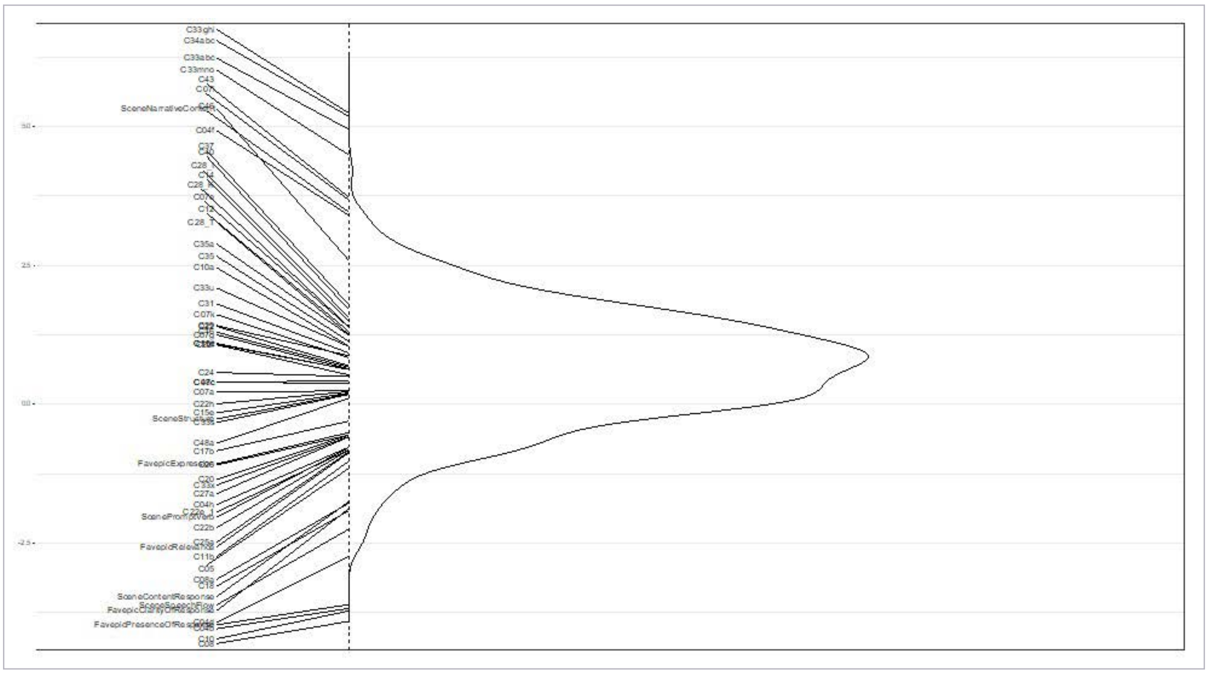
Figure 4 Plot of item delta weighted mean square fit statistics against quantiles of the standard normal distribution



Scale reliability is very high. Plausible value (PV) reliability is 0.97 and weighted likelihood estimation (WLE) separation reliability is 0.94. Note that with more than 100 polytomous items in the test (summing to more than 200 category boundaries) administered to more than 500 children on two time points, a high reliability is likely for most models. For example, PV reliability at any one time point, is 0.79 in 2018 and 0.88 in 2019.

Targeting is good, although there are a number of items outside the range of the ability distribution (the density shown on the right of Figure 5). There are also relatively fewer items in the upper range of the ability distribution and this is something to be better targeted in future item development.

Figure 5 Wright Map of item location (difficulties) relative to the distribution of children's abilities

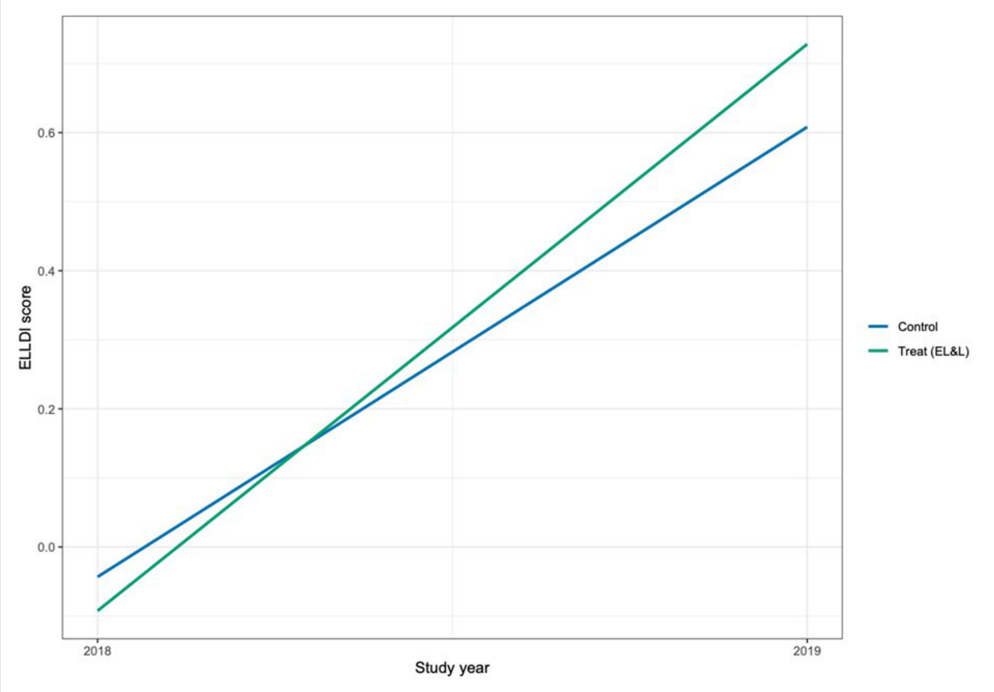


Note only a selection of item labels are printed to minimise overlap.

Scaling and modelling

Holding item parameters constant at both time points, the average growth between 2018 and 2019 is 0.70 logits (approximately 0.63 standard deviations [SD] of the 2019 ability distribution). The correlation between children's scores at time 1 and 2 is 0.78. In an unconditional regression model, there is a small positive effect for the interventions. Children in the intervention group start marginally behind the control group (ES = –0.05, *ns*) and have grown slightly faster to be, on average, above the control group at the end of time 2 (0.14 SD, $p < 0.05$), as shown in Figure 6. Caution should be made with interpreting these interim results as the longitudinal design is not complete, with a third wave of data collection completed in 2021.

Figure 6 Average linear growth trajectories for study children between time 1 and 2



Discussion

This paper shows that it is possible to develop an assessment that aligns both with guiding principles of good quality assessment in the yearly years and cutting-edge psychometric techniques. The development of assessments that access a developmental continuum (and describe them – so called *learning progressions*) is growing quickly in school education circles. We show that this kind of progress can be made in ECEC and can be done in a way that is familiar to educators: one-on-one interactions with children around activities like reading a picture book.

The utility of strong measurement and description of oral language and pre-literacy development is clear. If educators can see, on a continuum, 1) where a child is currently performing in terms of their oral language and pre-literacy skills and 2) what are the skills and abilities that typically develop next, then they can focus their efforts on making conceptually sound incremental improvements that are targeted as individual child needs. This approach will also build educator capacity by providing a conceptual understanding of the construct they are supporting children's growth in.

Ongoing work is required to develop an item bank that covers the range of abilities that cover the typical oral language and literacy development of children aged 2 to 8 years. Further, it is important to carefully assess the calibration of this measure as more data is collected. Future work should include moving towards adaptive assessment or at least providing a way for educators to ensure the tasks and items given to children are optimally targeted at their level of ability. The development of automatic scoring and reporting will also ensure educators get access to high-quality learning data on the same day they complete an assessment of a child.

References

Adams, R. J., & Wu, M. (2009). The construction and implementation of user-defined fit tests for use with marginal maximum likelihood estimation and generalized item response models. *Journal of Applied Measurement*, *10*(4).

Adams, R. J., Wu, M. L., Cloney, D., & Wilson, M. R. (2020). *ACER ConQuest: Generalised Item Response Modelling Software* (Version 5) [Computer software]. Australian Council for Educational Research.

Anzai, D., Knowles, S., Cloney, D., Munro-Smith, P. & Mitchell, P. (2021). *Assessment of oral language and early literacy in early childhood education and care: Literature review*. [Manuscript submitted for publication]. Victorian Curriculum and Assessment Authority.

Australian Institute for Teaching and School Leadership. (2020). *Reading instruction evidence guide*. AITSL. https://www.aitsl.edu.au/docs/default-source/initial-teacher-education-resources/reading-instruction/reading-instruction-evidence-guide.pdf?sfvrsn=5465d73c_2

Cloney, D., & Adams, R. J. (2021). *conquestr* (Version 0.9.9) [Computer software]. Australian Council for Educational Research. https://cran.r-project.org/package=conquestr

Cloney, D., Jackson, J., & Mitchell, P. (2020). *Assessment of children as confident and involved learners in early childhood education and care: Literature review*. Victorian Curriculum and Assessment Authority. https://www.vcaa.vic.edu.au/Documents/earlyyears/EYLitReviewLearning.pdf

Foorman, B. R., Herrera, S., Petscher, Y., Mitchell, A., & Truckenmiller, A. (2015). The structure of oral language and reading and their relation to comprehend in Kindergarten through Grade 2. *Reading and Writing*, *28*(5), 655–681. https://doi.org/10.1007/s11145-015-9544-5

Konza, D. (2014). Teaching reading: Why the 'Fab Five' should be the 'Big Six'. *Australian Journal of Teacher Education*, *39*(12), 153–169.

Linacre, J. M. (1993). *Generalizability theory and many-facet Rasch measurement.* American Educational Research Association.

National Institute of Child Health and Development. (2000). *Report of the National Reading Pane. [Teaching children to read] An evidence-based assessment of the scientific research literature on reading and its implications for reading instructions*. (No. 00-4769). Washington DC: Government Printing Office. http://www.nichd.nih.gov/publications/nrp/smallbook.cfm

Palermo, F., Hanish, L., Martin, C., Fabes, R., & Reiser, M. (2007). Preschoolers' academic readiness: What role does the teacher–child relationship play?, *Early Childhood Research Quarterly*, *22*(4), 407–22. https://doi.org/10.1016/j.ecresq.2007.04.002.

Pianta, R. C., La Paro, K. M., & Hamre, B. K. (2008). *Classroom Assessment Scoring System: Manual K–3*. Brookes Publishing Co.

R Core Team (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. www.R-project.org/

Reynolds, B. (2020). Assessment and documentation for children's learning and development. In Kilderry, A. & Raban, B. (Eds.), *Strong foundations: Evidence informing practice in early childhood education and care* (pp. 250–265). ACER Press.

Rowe, K. (2005). *Teaching reading: National inquiry into the teaching of literacy*. Australian Government Department of Education, Science and Training. https://research.acer.edu.au/tll_misc/5

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI Monograph Series*, *2*, 9–36.

Wu, M. (2005). The role of plausible values in large-scale surveys. *Studies in Educational Evaluation*, *31*(2–3), 114–128. https://doi.org/10.1016/j.stueduc.2005.05.005

This time without 'feeling': Children's intuitive theories of art as a logical basis for learning progression in visual arts

Dr Karen Maras
UNSW Sydney

https://doi.org/10.37517/978-1-74286-638-3_3

Dr Karen Maras is a senior lecturer in the School of Education at the University of New South Wales. As a specialist in visual arts education, she is Program Convenor, Visual Arts, Media Arts and Design Education. Prior to this appointment Karen led programs in visual arts teacher education at Australian Catholic University and also held a range of leadership roles in the ACU Faculty of Education NSW. She began her career in academia at COFA UNSW wherein she completed her undergraduate, postgraduate and doctoral studies. Karen's contributions to visual arts teacher education at UNSW Sydney and curriculum development in state and national contexts are complemented by her research on students' conceptual development, learning and curriculum issues in visual arts.

Abstract

Learning in Visual Arts has traditionally been framed as an experiential process in which feeling and intuition complement the development of aesthetic knowledge. However, while art can be about feelings and processes that develop students' expressive capacities, the complexity of art understanding and thinking extends beyond this narrow common-sense assumption. I argue that this assumption, which is represented in the *Australian Curriculum: The Arts* (ACARA, 2015), and even more firmly resonates in recent proposals for the revision of this curriculum (ACARA, 2021), obfuscates the conceptual and theoretical bases on which students make progress in art understanding. This paper examines the proposition that art understanding emerges progressively and can be described in conceptual terms, the basis of which can be identified in empirical research on the emergence of children's intuitive theories of art. This paper examines how selected studies articulate the cognitive grounds on which students' ontologies of art and epistemological beliefs are represented in their reasoning about art over time. It is argued that an empirically supported conception of learning anchored in students' cognitive development in art that recognises the theoretical commitments underscoring their conceptual and practical reasoning in visual arts practices K–12 provides a logical basis for articulating progression in the subject.

Introduction

The question of how learning progression is described in Visual Arts remains a vexed issue in current curriculum developments in Australia. Ongoing debates about learning progression feature in discussions of the *Australian Curriculum: The Arts F–10* (Australian Curriculum, Assessment and Reporting Authority [ACARA], 2015) (AC:TA) and the Visual Arts curriculum in New South Wales (NSW), both of which are the focus of concurrent and intertwined curriculum reviews (ACARA, 2020; New South Wales Educational Standards Authority [NESA], 2020).

The AC:TA starts from the position that process-based learning in art will yield embodied understandings of aesthetic knowledge. Through making and responding, students engage aesthetic knowledge (e.g. skills, techniques, conventions, elements, materials). Advocates of this approach argue that process-based learning provides the means for realising self-expressive ends. This curriculum starts from the premise that all students are intuitively pre-disposed to be aestheticians and will develop increasing sophistication through emotive enagement and aesthetic experience. By contrast, advocates for the approach taken in the NSW Visual Arts curriculum K–12 argue that process fails to make explicit the role of thinking about art and how this informs what students understand as makers and critics (Brown, 2017). The NSW curriculum starts from the position that students learn practically and conceptually through reasoning in the practices of art making, art criticism and art history (Board of Studies New South Wales, 2003). Engagement with core concepts, explanatory systems and a theory of practice support them to demonstrate increasing intellectual autonomy as they develop understandings of the relationship of conceptual knowledge and practical activities when making art or constructing critical interpretations (Maras, 2021). This conception of learning in art is grounded in empirical research focusing on the students' developing theories of art.

In this paper I argue that understanding how students intuitively understand art at different points in their development provides a basis on which progression in art learning can be described. To explore this issue of learning progression in visual arts, discussion will 1) briefly define learning progression, and 2) examine selected cross-age studies which illustrate how changes in students' intuitive theories of art change over time and provide a basis for understanding how students learn in visual arts. The paper concludes with some reflections on the terms on which research on students' intuitive theories of art provides a logical basis on which a high quality, academically rigorous curriculum F–12 can be developed.

Learning progression

Grounded in empirical studies of cognitive development, learning progression describes learning growth, placing 'explicit emphasis on ways students' thinking becomes more sophisticated over time in terms of interactions between their growing understanding of content … and their ability to use that understanding in reasoning…' (Mosher, 2011, p. 3) in domain-specific terms. Drawing from Driver et al. (1994), the development of learning progression concerns identifying three factors; 'changes in students' ontologies within specific domains, changes in reasoning strategies, and changes in epistemological commitments', a project that involves mapping the conceptual development of younger and older students (Duschl, 2019, p. 97). As such, learning progression provides a map of iterative changes in students' understandings over time, a map that can support teachers in better anticipating the support individual students need to reach particular goals, such as curriculum standards over the duration of their learning in a specific knowledge domain (Masters, 2013). The following discussion outlines selected studies that have contributed to the stock of knowledge about conceptual development in art.

Empirical research on children's conceptual development in art

Empirical studies of students' conceptual development in art have revealed progressive changes in their ontologies of art, their reasoning and the epistemological commitments during early childhood through early adolescence. Studies of this kind focus on reasoning performances by students about the nature and function of artworks as things that exist in the world. For example, the theoretical foundations of students' intuitive theories of art have been identified (Freeman, 1995). By mapping conceptual patterns in students' general reasoning about artworks, this study identified the ontological terms on which younger and older students formulated reasons for how and why artworks exist in relation to artists, the world and beholders. Answers to questions devised to test students' thinking about these relationships revealed that younger students hold naïve realist theories of what artworks are as things in the world, equating subject matter with artwork function.

Gradually this naïve view develops into a more sophisticated notion of artworks as representations of subject matter made by artists who draw feelings, desires and motives to do so. With the advent of intentional claims about an artwork's existence, older students progress their ideas to artworks as intentional artefacts, describing their function in the world in relation to artists as producers of art who invest artworks with intentional value in anticipation of thoughtful reception by beholders.

On these terms, students' developing ontologies of art are complemented by the development of reasoning skills when construing and justifying relationships among the concepts of artwork, beholder, subject matter and artist in which a theory of mind is demonstrated. These developments were also underscored by students' capacities to consult and apply intentional beliefs to argue how artworks were situated within intentional relationships between artworld concepts, an advance which also included a growing awareness of their own intentional agency as beholders. The conceptual and theoretical trajectory described in this example has also been confirmed by subsequent studies of students' general reasoning about art (Brown & Freeman, 1993; Freeman & Sanger, 1993, 1995). However, perhaps a consequence of engaging students in general conceptual reasoning, this series of studies on students' intuitive theories of artworks did not detect any evidence that children are innately orientated to aesthetic beliefs as epistemological grounds for talking about art.

Research by Maras (2010) explored in greater depth the theoretical and practical bases of children's critical reasoning in art. This study of younger and older students' critical judgements of artworks explored the terms on which they 'recognise and identify' (Wollheim, 2001) the meaning and value of examples of artworks they judged to be 'good'. While engaged in a curatorial task in which they were asked to recommend good artworks to the researcher for inclusion in an exhibition, students were invited to reason out their choices of artworks. Analysis of their reasoning revealed a great deal about changes in their understanding of art, the ontological bases on which they represent their claims about the function of artworks, and the kinds of epistemological beliefs they bring to bear in their judgements of artworks. With age, students gain greater control of their reasoning skills, mastering recursive transitions in which initial ideas are gradually renovated as they construct increasingly more complex, higher order claims about artwork meaning and value. As their dexterity in reasoning skills increases, so does the scope of the ontological bases on which they recognise artworks' properties and identify them on intentional terms. This shift represents a transition from naïve realism at the age of six to a reasonably replete account of artworks and their properties as products of artists' intentions and beliefs by the age of nine. Some students aged nine could also extend their judgements to include consideration of the role of the audience in their explanations of representational relations. Advances into intentional ontologies art are then consolidated by 12 years of age, a development confirmed by students' developing skills in consulting a range of epistemological beliefs. At this age students were well on the way to ascribing artworks' meaning and value as cultural status symbols, forms of intuitive expression and representations of style and taste. In other words, students begin to locate the function of artworks as functions of intentional transactions of the kind that occur in the art world as a social reality.

The conceptual changes identified in these two examples of empirical research on students' art understanding during the primary and early secondary years of schooling describe a pathway of understanding from ontologically naïve assumptions about how artworks exist to highly complex and robust intentional theories of art. Their intuitive assumptions about art reflect a conceptual logic and sequential consistency that is complemented by advancing skills in controlling their reasoning to represent and justify points of view. This pathway also reflects students' advance into understanding the intentional function of artworks as artefacts that are shaped by beliefs that artists adopt when producing them and that are shared among audiences of art. This trajectory in students' art understanding provides a logical basis, or a set of logical constraints on which descriptions of learning progression in visual arts can be developed (Maras, 2018).

Conclusion

This brief account of research on students' intuitive theories of art reveals how art understanding is informed by the practicalities of building reasoned critical accounts of art that are anchored in ontological commitments and changing epistemological orientations (Driver, et al., 1994). I have argued that insight into how thinking informs skilful activity (i.e. practical and conceptual reasoning) is central to mapping students' progression in art, a factor that also supports understanding of how to teach students with a view to deepening understanding of the core concepts and principles of learning in art in time and over time (NESA, 2020).

The account of learning progression outlined in this paper places emphasis on the 'development of mind' in visual arts (Eisner, 2003). The scope of conceptual and practical reasoning as a basis for learning progression in art moves beyond intuitive processes and aesthetic knowledge to embrace a rich form of conceptual engagement in the subject that compliments the practicalities of skills, capabilities and processes in art practice (Brown, 2017). Despite pressure to 'adopt and adapt' the AC:TA in the NSW context, visual arts educators in NSW have rejected the generic framing of learning as process in AC:TA. They remain committed to retaining their current curriculum, arguing that an empirically supported conception of learning grounded in students' development in art that recognises the theoretical commitments underscoring their conceptual and practical reasoning in visual arts practices K–12 provides a logical basis for articulating progression in the subject. This approach to establishing a basis for teaching and assessing for depth in understanding, as well as establishing a basis on which to develop progressions has already been laid in the current curriculum provisions in NSW (Maras, 2021). However, a great deal more empirical research on student's reasoning in both art making and art interpretation the middle years of schooling would support a well-rounded understanding of students' intuitive theories of art as a basis describing learning progression and for engaging in good teaching.

References

Australian Curriculum, Assessment and Reporting Authority. (2015). *The Australian Curriculum: The Arts*. https://www.australiancurriculum.edu.au/f-10-curriculum/the-arts/.

Australian Curriculum, Assessment and Reporting Authority. (2020). *Terms of reference – review of the Australian Curriculum F–10*. https://www.acara.edu.au/docs/default-source/curriculum/ac-review_terms-of-reference_website.pdf.

Australian Curriculum, Assessment and Reporting Authority. (2021). *The Arts Consultation Curriculum: The Australian Curriculum Review*. https://www.australiancurriculum.edu.au/consultation/the-arts/

Board of Studies New South Wales. (2003). *Years 7–10 visual arts syllabus*. https://www.boardofstudies.nsw.edu.au/syllabus_sc/pdf_doc/visual_arts_710_syl.pdf

Brown, N. C. M. (2017). *Studies in philosophical realism in art, design and education*. Springer International Publishing.

Brown, N. M., & Freeman, N. H. (1993). *Children's developing beliefs as a basis for sequencing in art education*. Australian Association for Research in Education. https://www.aare.edu.au/data/publications/1993/brown93025.pdf

Driver, R., Leach, J., Scott, P., & Wood-Robinson, C. (1994). Young people's understanding of science concepts: Implications of cross-age studies for curriculum planning. *Studies in Science Education*, *24*(1), 75–100. https://doi.org/10.1080/03057269408560040

Duschl, R. A. (2019). Learning progressions: Framing and designing coherent sequences for STEM education. *Disciplinary and transdisciplinary Science Education Research*, *1*(1), 1–10. https://doi.org/10.1186/s43031-019-0005-x

Eisner, E. W. (2003). *The arts and the creation of mind*. Yale University Press.

Freeman, N. H. (1995). The emergence of a framework theory of pictorial reasoning. In C. Lange-Kuttner & G. V. Thomas (Eds.), *Drawing and looking: Theoretical approaches to pictorial representation in children* (pp. 135–146). Harvester Wheatsheaf.

Freeman, N. H., & Sanger, D. (1993). Language and belief in critical thinking: Emerging explanations of pictures. *Exceptionality Education Canada*, *3*(3), 43–58.

Freeman, N. H., & Sanger, D. (1995). Commonsense aesthetics of rural children. *Visual Arts Research*, *21*(2), 1–10.

Maras, K. (2010). Age-related shifts in the theoretical constraints underlying children's critical reasoning in art. *Australian Art Education*, *33*(1), 20–28. https://search.informit.org/doi/10.3316/informit.431497643821282

Maras, K. (2018). Mind, language and artworks as real constraints on students' critical reasoning about meaning in art. *The International Journal of Art & Design Education*, *37*(3), 530–540. https://doi.org/10.1111/jade.12152

Maras, K. (2021). Ahead of its time? Reassessing what is core content in Visual Arts in the New South Wales Curriculum. *Curriculum Perspectives*, 1–11. https://doi.org/10.1007/s41297-021-00136-6

Masters, G. (2013). *Reforming education assessment: Imperatives, principles and challenges* (Australian Education Review No. 57). http://research.acer.edu.au/aer/12/

Mosher, A. (2011). The role of learning progressions in standards-based education reform. *CPRE Policy Briefs, RB-52.* https://doi.org/10.12698/cpre.2011.rb52

New South Wales Educational Standards Authority. (2018). *Terms of Reference – NSW Curriculum Review*. https://nswcurriculumreform.nesa.nsw.edu.au/assets/docs/nsw-curriculum-review-terms-of-reference.pdf

New South Wales Educational Standards Authority. (2020). *Nurturing wonder and igniting passion: Designs for a new school curriculu*m. https://nswcurriculumreview.nesa.nsw.edu.au/home/siteAreaContent/524abec1-f0f9-4ffd-9e01-2cc89432ad52.

Wollheim, R. (2001). On aesthetics: A review and some revisions. *Literature and Aesthetics*, *11*, 7–32.

Learning progressions as models and tools for supporting classroom assessment

Associate Professor Alicia C Alonzo
Michigan State University

https://doi.org/10.37517/978-1-74286-638-3_5

Alicia Alonzo, PhD, is an Associate Professor in the Department of Teacher Education at Michigan State University. Her research agenda is centred around the premise that assessment – broadly construed – has the potential to exert a significant influence on both student and teacher learning. Her work considers assessment practices at both the classroom level (as teachers use interactions with students to tailor instruction and generate professional knowledge) and at the large-scale level (as state, national, and international assessments signal what is valued as learning in science classrooms). Much of Alicia's research has focused on tools (learning progressions) and knowledge (pedagogical content knowledge) underlying classroom assessment practices.

Abstract

Like all models, learning progressions (LPs) provide simplified representations of complex phenomena. One key simplification is the characterisation of student thinking in terms of levels. This characterisation is both essential for large-scale applications, such as informing standards, but potentially problematic for smaller-scale applications. In this paper, I describe a program of research designed to explore the smaller-scale use of LPs as supports for teacher classroom assessment practices in light of this simplification. Based on this research, I conclude that LP levels may serve as a generative heuristic, particularly when teachers are engaged with evidence of the limitations of LP levels and supported to use LPs in ways that are not reliant on these levels.

Introduction

Like all models, LPs provide simplified representations of a complex phenomenon (e.g. Lehrer & Schauble, 2015), capturing some features of students' thinking and learning but necessarily simplifying others. As Leher and Schauble (2015) caution:

> It is imperative to remember that an LP is a model … Like all models … LPs are incomplete and even incorrect in some respects … The question to ask about [LPs] is not 'Are they true?' but rather, 'Are they useful for the purposes that we need them to achieve?' (p. 435).

In other words, as with scientific models, LPs should be evaluated not just in terms of empirical adequacy but also in terms of criteria such as utility and generativity (Odenbaugh, 2005) with respect to proposed use.

The purposes proposed for LPs include both large-scale applications – informing standards (Foster & Wiser, 2012), curricula (Songer et al., 2009), and large-scale assessment (Alonzo et al., 2012)—and those at a smaller-scale – informing instruction (Scott et al., 2019) and classroom assessment (Furtak, 2012). While large-scale uses rely on the broad characterisation of student thinking into LP levels, this simplification may be problematic for smaller-scale uses of LPs (Alonzo & Elby, 2019; Alonzo et al., 2021). In particular, what underlies common approaches to validation is the assumption that student thinking is coherent and consistent. LPs are evaluated according to the criterion of 'conceptual coherenc[e]' (Anderson, 2008, p. 4) and using assessments that rely on the assumption that student thinking is consistent enough to be reliably characterised using LP levels. However, research suggests that student ideas have context-dependencies that do not fit neatly into LP levels (Heredia et al., 2012) and that the conceptual territory between the upper and lower anchors of an LP (the 'messy middle'; Gotwals & Songer, 2010, p. 277) may be particularly fragmented and context-dependent (Steedle & Shavelson, 2009).

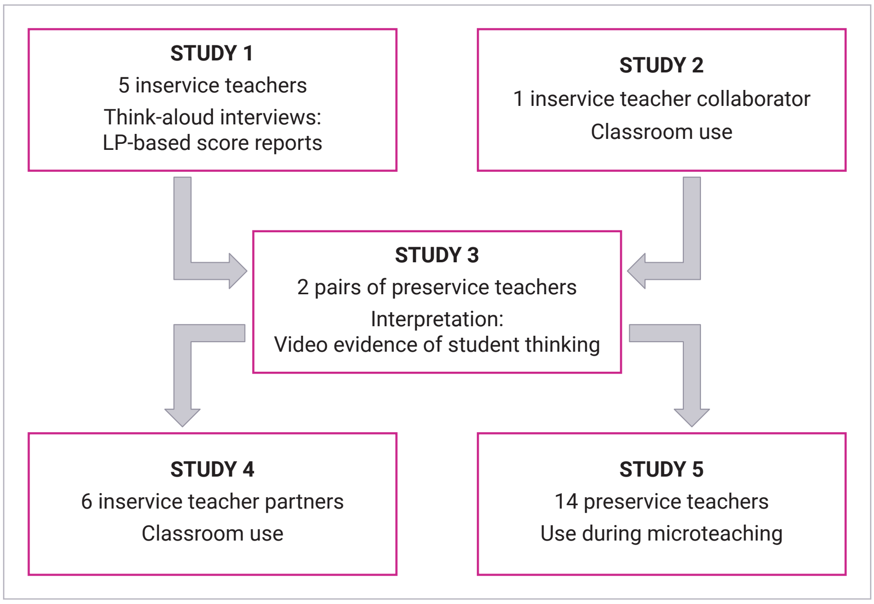
To support classroom assessment, LPs have been proposed to focus attention on student ideas that might be important to elicit, support interpretation of the ideas that are elicited, and inform responses to those ideas (Alonzo, 2018). A basic model for this use focuses on determining students' LP levels and using those levels to identify instructional next steps that will support incremental movement towards the targeted level of understanding (Covitt et al., 2018). However, diagnosing and responding based on LP levels is complicated by the uncertainty (and, thus, unreliability) introduced by the context-dependency and inconsistency of students' thinking.

In this paper, I describe a series of studies designed to support teacher use of LPs to support classroom assessment practices, with particular attention to a) how LPs might have utility and generativity for this purpose despite the simplification of LP levels and b) how LPs might be used without relying on this simplification.

Research program

In order to explore the usefulness of LPs as supports for teachers' classroom assessment practices, particularly in light of the simplified representation of student thinking in levels, we undertook a series of five related studies (depicted in Figure 1). The first four are discussed.

Figure 1 Relationship between five studies comprising research program on teacher use of LPs to support classroom assessment practices



Study 1: Inservice teachers' engagement with LP-based score reports

Study 1 (Alonzo & Elby, 2019) involved five experienced high school physics teachers with high-quality formative assessment practices (based on researcher recommendation and verified using an initial interview). The teachers interacted with an LP-based score report in a set of two think-aloud interviews. The score report presented LP-based assessment results for a fictitious, but realistic, class of students. We examined: a) the assumptions about student thinking that teachers made as they engaged with the score reports and the instructional reasoning supported by different assumptions and b) ways that teachers developed (or could develop) new understandings based on their engagement with the score report.

We found that teachers understood the general intent of the LPs and appropriated language from the LP-based materials to reason about student thinking. However, they more frequently treated student thinking as less coherent than the LP perspective suggested and offered finer-grained interpretations than those provided by LP levels. The only specific, actionable instructional responses that teachers proposed were based on these finer-grained analyses. Teachers generated knowledge about student thinking by taking a sceptical stance towards the LP and investigating places where the LP model did not adequately explain the data.

Study 2: Inservice teacher collaborator's use of LPs

Study 2 (Alonzo, 2019; Alonzo & Elby, 2019) focused on one teacher from Study 1 who collaborated with (and participated as member of) the research team. 'Tim' used the LPs to inform his classroom assessment practices, serving as a check of the laboratory-based results obtained in Study 1.

Tim used the LP levels to fulfill a requirement that he report pre-/post-test gains for his students, highlighting that level-based interpretations permitted him to demonstrate growth even when students had not met the learning targets. However, he did not find the levels useful for informing his own classroom assessment practices; instead, as in Study 1, he used LP-based assessment results as a springboard for inquiry into his students' ideas.

Study 3: Preservice teachers' use of LPs to interpret video evidence of student thinking

In Study 3 (von Aufschnaiter & Alonzo, 2018), we added a short introduction (150 minutes) to a LP to a methods course for preservice physics teachers. Using a pre- and post-design, we compared how two pairs of teachers interpreted evidence of student thinking presented in videos of cognitive interviews with and without the LP.

We found that the LP seemed to support the teachers in attending to specific aspects of students' thinking highlighted in the LP, in avoiding speculative or unwarranted interpretations, and in thinking about implications for future learning in more student-centred ways. At the same time, we cautioned that preservice teachers may over-apply LP frameworks – for example, attending only to ideas represented in the LP or assuming that students who hold one idea at a given LP level hold all ideas at that level.

Study 4: Inservice teachers' use of LPs to support classroom assessment practices

In Study 4 (Alonzo et al., 2021), we supported six high school physics/physical science teachers over two years to incorporate LPs into their classroom assessment practices. Through two summer workshops and support via planning meetings during the ensuing academic years, we introduced the LP model (and three specific LPs) and highlighted ways that student thinking is more complex than is represented in LP levels. Instead of prescribing how we thought teachers should use LPs, we encouraged teachers to take up these tools in ways that made sense to them.

We found that teachers' LP use varied – from completely reliant on LP levels (e.g. reporting student performance using LP levels) to not using LP levels at all (e.g. attending to ideas on the LP without attention to levels). However, the uses that were less reliant on LP levels were more prevalent, and teachers described challenges in using LP levels – for example, students' held ideas at different LP levels, which complicated level-based interpretations. While some of the teachers' challenges may be related to models of thinking and learning inconsistent with the more constructivist assumptions of the LPs, we concluded that teacher use of LPs represented a rational response to level-based challenges.

Conclusion

Although teachers in our studies used LP levels to elicit, interpret, and respond to student ideas, these uses appeared challenging and were not prevalent. Teachers more often – and seemingly more productively – used LPs in ways that considered student ideas more closely (e.g. attended to individual ideas within an LP level, rather than the level as a coherent whole) and/or did not attend to LP level. We viewed this as a reflection of the potential issues with LP levels, particularly in relation to the inconsistency of student thinking that teachers observed (both in their own students and in the evidence we provided to them).

At the same time, the basic model of diagnosing students' LP levels and using those diagnoses to identify appropriate instruction seems to serve as a useful heuristic and jumping off point for teacher investigations of student thinking and of ways LPs might be useful to them. In this way, LPs (with their levels) appear to be a productive site for teachers' generation of knowledge-*of*-practice (Cochran-Smith & Lytle, 1999). At the same time, teachers used LPs in ways that did not rely on LP levels but that did reflect an 'LP approach' – focusing on student ideas and how they can be supported to change gradually over time. Especially because researcher-developed LPs do not exist for all topics in the K–12 curriculum, this more general approach may be an important way that LPs can impact instructional practice even if their levels do not support strict diagnoses and prescriptions for instructional 'next steps.'

References

Alonzo, A. C. (2018). An argument for formative assessment with science learning progressions. *Applied Measurement in Education*, *31*, 104–112.

Alonzo, A. C. (2019). Defining trustworthiness for teachers' multiple uses of classroom assessment results. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 120–145). Routledge.

Alonzo, A. C., & Elby, A. (2019). Beyond empirical adequacy: Learning progressions as models and their value for teachers. *Cognition and Instruction*, *37*, 1–37.

Alonzo, A. C., Neidorf, T., & Anderson, C. W. (2012). Using learning progressions to inform large-scale assessment. In A. C. Alonzo & A. W. Gotwals (Eds.), *Learning progressions in science: Current challenges and future directions* (pp. 211–240). Sense Publishers.

Alonzo, A. C., Wooten, M. M., & Christensen, J. A. (2021). *Teachers' perceived uses of learning progressions to inform classroom assessment and related practices* [Manuscript submitted for publication].

Anderson, C. W. (2008, February). *Conceptual and empirical validation of learning progressions*. Consortium for Policy Research in Education. http://www.cpre.org/ccii/images/stories/ccii_pdfs/learning%20progressions%20anderson.pdf

Cochran-Smith, M., & Lytle, S. L. (1999). Relationships of knowledge and practice: Teacher learning in communities. *Review of Research in Education*, *24*, 249–305.

Covitt, B., Gunckel, K. L., Syswerda, S., & Caplan, B. (2018). Teachers' use of learning progression-based formative assessment in water instruction. *Applied Measurement in Education*, *31*, 128–142.

Foster, J., & Wiser, M. (2012). The potential of learning progression research to inform the design of state science standards. In A. C. Alonzo, & A. W. Gotwals (Eds.), *Learning progressions in science: In science: Current challenges and future directions* (pp. 435–460). Sense Publishers.

Furtak, E. M. (2012). Linking a learning progression for natural selection to teachers' enactment of formative assessment. *Journal of Research in Science Teaching*, *49*, 1181–1210.

Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education*, *94*, 259–281.

Heredia, S., Furtak, E. M., & Morrison, D. (2012, March 25–28). *Item context: How organisms used to frame natural selection items influence student response choices.* [Paper presentation]. NARST annual international conference, Indianapolis, IN, United States.

Lehrer, R., & Schauble, L. (2015). Learning progressions: The whole world is NOT a stage. *Science Education*, *99*, 432–437.

Odenbaugh, J. (2005). Idealized, inaccurate, but successful: A pragmatic approach to evaluating models in theoretical ecology. *Biology and Philosophy*, 20, 231–255.

Scott, E. M., Wenderoth, M. P., & Doherty, J. H. (2019). Learning progressions: An empirically grounded, learner-centered framework to guide biology instruction. *CBE—Life Sciences Education*, *18*, 1–11.

Songer, N. B., Kelcey, B., & Gotwals, A. W. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching*, *46*, 610–631.

Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching*, *46*, 699–715.

von Aufschnaiter, C., & Alonzo, A. C. (2018). Foundations of formative assessment: Introducing a learning progression to guide preservice physics teachers' video-based interpretation of student thinking. *Applied Measurement in Education*, *31*, 113–127.

Applying empirical learning progressions for a holistic approach to evidence-based education: SWANs/ABLES

Dr Emily White
Melbourne Graduate School of Education

https://doi.org/10.37517/978-1-74286-638-3_6

Dr Emily White is a research fellow and sessional lecturer at the Melbourne Graduate School of Education, and a teacher practitioner with over 15 years' experience in inclusive and disability-specific education. Her research interests include accessible and valid assessment for students with disability, digital literacy, and understanding and supporting the learning of students with vision impairment using teacher expertise and evidence. She is currently working on an Australian Research Council Discovery Grant project exploring self-regulated learning in STEM classes. Emily lectures on using data to inform teaching for students with disability, and supporting access to learning via technology. Her PhD study investigated the development and validation of measures for assessing and teaching digital literacy for students with disability. Her assessment tool, derived empirical learning progression, and matched teaching strategies are now used in over 400 schools across Australia and abroad. At the Statewide Vision Resource Centre in Melbourne, she teaches and provides professional learning, having previously served in curriculum coordination, specialist, itinerant, and classroom roles across a range of settings.

Abstract

Learning progressions have become an increasing topic of interest for researchers, educational organisations and schools as they can describe the expected pathway of learning within a content area to allow for targeted teaching and learning at all levels of ability. However, there is substantial variation in how learning progressions are developed and to what extent teachers can use them to inform their practices. The ABLES/SWANs tools (Students with Additional Needs/Abilities Based Learning and Education Support) are an example of how an empirical learning progression can be applied to support teachers' ability to not only target teaching to a student's Zone of Proximal Development (Vygotsky, 1978), but also to plan, assess, and report on learning. Across Australia, these tools are used to help of thousands of teachers of students with disability to make evidence-based teaching and learning decisions and demonstrate the impact of their work with students. This approach, which scaffolds student achievement towards goals informed by an empirical learning progression, combined with reflective teaching practices, can help teachers to develop their capacity as professionals and provide the most effective teaching and learning for every student, regardless of the presence of disability or additional learning need.

Introduction

Learning progressions are gaining a substantial foothold in the minds of researchers and the practices of teachers. Global organisations such as the International Bureau of Education of the United Nations Educational, Scientific and Cultural Organization (UNESCO) have become increasingly interested in developing and applying learning progressions to support the learning of students around the world (Marope et al., 2019). On a national level, the Australian Council for Educational Research (ACER), the Australian Curriculum, Assessment, and Reporting Authority (ACARA), and the Assessment Research Centre at the Melbourne Graduate School of Education are all working on building learning progressions using different methodologies.

While not all learning progressions are the same, they do follow the same premise: that the learning in a domain unfolds along a long-term developmental pathway, which can be described in increasing levels of difficulty and complexity (Heritage, 2008; Wiliam, 2007). This view contrasts with the notion that learning is simply the acquisition of age- or grade-level content to be taught. Rather, it positions the learner at the centre of a targeted teaching and learning process that addresses their current readiness to learn, in order to support the understanding of learning as a transformative process of increasingly sophisticated skill or understanding (White, 2019). By doing so, learning progressions 'represent not only how knowledge and understanding develops, but also predict how knowledge builds over time' (Stevens et al., 2002, p. 2). The provision of this critical information can support teachers to understand what a learner currently knows and can do in order for them to make good, evidence-based decisions about what a learner is ready to learn next and how best to support the learner towards that goal.

The idea that assessment might be 'as, for, and of' teaching has been strongly challenged by eminent scholars such as Patrick Griffin, who argue that assessment is solely for informing targeted teaching for every learner. The use of assessment to build a picture of a learner's current capabilities in a domain, as evidenced by what they can do, make, say or write, in conjunction with an empirical learning progression that describes the expected pathway of learning in that domain, can support the development of high-quality learning goals and planning to scaffold the learner along the pathway (Griffin, 2014). To understand the differences between empirical learning progressions and others, it is necessary to consider the methodology used to build a learning progression.

Empirical learning progressions

Empirical learning progressions refer to learning progressions developed using methodologies that incorporate data on student learning, as well as the knowledge of researchers and/or teachers. Not all methodologies for learning progressions use such an approach; some rely exclusively on the knowledge of researchers and/or teachers to describe a hypothesised learning progression in a domain such as numeracy or literacy. Whenever possible, the merging of an expertise basis with an evidence basis allows for the richest description of expected student learning in a domain, and can reflect the depth of knowledge held by expert teachers. For example, the outcomes of the Students with Additional Needs (SWANs)/Abilities Based Learning and Education Support (ABLES) project (SWANs/ABLES project) (University of Melbourne, 2017, 2018), which drew upon scholarly literature, teacher expertise, and student data to build empirical learning progressions for students with additional learning needs and/or disability, demonstrated that expert teachers were highly skilled at describing the likely learning pathways for these students. The close match found between their hypothesised rubric for a learning domain, and the rubric derived from student data, supported the notion that many expert teachers possess an accurate internalised learning progression created from many years of experience and high levels of training (see, for example, White, 2019; Woods, 2010).

The creation of an empirical learning progression requires substantial research and student data, and is thus more difficult and time-consuming to build than those built on a more hypothetical basis. As a result, many learning progressions are hypothetical in nature, though may still retain some strong arguments for validity. When considering a learning progression for use in schools, teachers should enquire about the methodological basis for the learning progression, and whether it was developed for the cohort of learners the teacher seeks to support. It is important that the span of capability described in a learning progression is inclusive of the range of domain-specific capabilities present in those learners; otherwise, the learning progression will not be valid for use with them.

Matching assessment with learning progressions

Another important factor to consider when seeking to use a learning progression is whether it has a matched assessment. A matched assessment tool that has been found to be valid and reliable for a cohort of students can allow teachers to more accurately locate a student within a level described in a learning progression, and can serve as an indicator that a learning progression is empirically derived. Without a matched assessment tool, it can be more difficult to determine the most appropriate level for a student, as most learners will have a degree of ability in some aspects of a domain that belongs in a higher level, but are less skilled in other aspects of the domain. Having a matched assessment also allows teachers to use the same assessment at different time points to more reliably judge a learner's growth within the learning progression, so to reflect on which teaching practices and other influences have supported – or hindered – progress. Last, the results of the matched assessment can be used for reporting purposes, so to celebrate success with the learner and their family, to demonstrate the impact of specific learning interventions, and discuss what new or different approaches might be taken to further support learning.

The SWANs/ABLES Project

An example of an empirical approach to developing learning progressions derived from teacher expertise and student data can be found in the SWANs/ABLES project (University of Melbourne, 2017, 2018) led by Patrick Griffin and Kerry Woods. This project created a holistic approach to assessment, planning, teaching, and reporting by using expert teacher knowledge to develop trial assessments that were used to collect data on the learning of thousands of students with disability and/or additional learning needs. The assessments were found to have strong arguments for their validity and reliability, so the data from them could be confidently used to build empirical learning progressions that described the likely trajectory of learning in nine foundational domains:

- literacy
- digital literacy
- numeracy
- communication
- social processes
- emotional understanding
- thinking skills
- learning skills
- movement.

Using teacher expertise and scholarly literature, teaching strategies were then developed to match each of the levels identified in each of the learning progressions, so to better support teachers to help learners progress from one level to the next. Lastly, multiple reporting formats were created to reflect a learner's location along the learning progression. Formats were devised for the purpose of reporting individual student ability and growth as well as class-wide and school-wide ability and growth over time as appropriate, both within and across domains. These formats allowed teachers and schools to celebrate student achievement with students and families based on evidence. They also allowed teachers and schools to reflect on differences in growth within an individual, and between groups of students, so to better understand the impact of planning and teaching decisions, resource provision, and external circumstances such as student absences. Today, hundreds of schools across Australia use SWANs or ABLES (a Victoria-specific version linked to the Victorian Curriculum) to support the achievement of students with disability and/or additional learning needs.

Assessment: Building a picture using evidence

Returning to Griffin's (2014) assertion that assessment is for teaching, the complementary statement is that assessment is for understanding. When a teacher seeks to understand what a student knows or can do, another way to conceptualise this is to understand the amount of current capability a student possesses in a particular domain, or, the amount of a latent trait that the student has. As a *latent trait* is not directly measurable, in the way that height or speed are, it must be assessed using indirect measures, such as by determining what a student can do, make, say, or write (Griffin, 2014) as evidence of how much capability they have in a skill, knowledge, or behaviour within a domain. By collecting evidence of what a student can do, make, say, or write, teachers can assemble a rich understanding of how much current capability a student has in a domain, for the purpose of informing targeted planning for that student's learning. Evidence of current capabilities could include photos or videos of student learning behaviours, samples of student work, and statements a student makes in an interview, for example. These are all examples of *observational assessment*, as teachers are using their observations of a student to inform their understanding of student capability. By comparing a student's capabilities against a high-quality, preferably empirical, learning progression, teachers can then make decisions about what the student is likely to be ready to learn next, as the teacher can compare current student capability with descriptions of the same skill/knowledge/behaviour at a more sophisticated or complex level.

As an example of using observational assessment to understand a student's current capabilities for the purpose of informing good planning and teaching, the SWANs/ABLES program allows teachers to use an online observational assessment tool to estimate a student's current location along the learning progressions for nine foundational learning domains. This program was developed to support teachers working with students identified as working below age- or grade-level for reasons of disability and/or additional learning needs. The observational assessment tool serves as a questionnaire version of a rubric that was developed using expert teacher knowledge, scholarly discourse, and student data, and found to have very strong arguments for reliability and multiple forms of validity. A teacher selects a learning domain, and, through a series of multiple-choice questions, draws upon their knowledge of the student to select the option for each question that best reflects what their student can do, make, say, or write on a typical day. The input from the teacher is then calculated using statistical analysis, and a description of the student's likely location on the learning progression is then displayed on a downloadable report, which also contains information about the levels above and below the student.

Understanding a student's location on the learning progression

A student's location on the learning progression can be understood as an estimate of their capability, as the true amount of a latent trait possessed by a student can never be fully understood or measured due to the indirectness with which it must be measured. On a more practical level for teachers, a student's capabilities will rarely be located entirely within one level, as it is expected that students possess different degrees of capability within the various skills, knowledges, and behaviours within a domain. So, while a student will be identified as generally working within a certain level though the assessment tool, teachers will often notice that a student is working at a slightly higher or lower level for some skills/knowledges/behaviours. As a result, teachers are encouraged to carefully read the student's identified level, as well as the levels above and below, in order to make the best decisions about what a student is ready to learn next – their Zone of Proximal Development (ZPD) (Vygotsky, 1978). By understanding each student's ZPD, teachers can begin to plan for individualised learning that supports students to meet learning goals based on what they are ready to learn next, rather than what an age- or grade-based curriculum says they should be taught next.

Setting targeted, evidence-based learning goals

Once teachers have identified a student's current individual level of skills, knowledges, and behaviours within a domain's learning progression, the next step is to find descriptions of the same skills, knowledges, and behaviours at a higher degree of sophistication or complexity. It may be that some students who are identified as working within the middle of a level (so, they have achieved approximately half of the learning expected in that level) should then focus on consolidating the remainder of that learning, which can form the basis for their learning goals. Other students who might be at the very beginning or end of a learning progression level may instead need to focus on some of the learning within another level to support them to move more fully into the next level. Regardless of a student's current location, it is necessary that teachers can support their decisions about what the student should learn next through the accumulation of evidence that indicates their current capability. By considering how a student's current skill, knowledge, or behaviour is expressed in a more sophisticated or complex manner, as well as the student's priorities, teachers can then begin to make decisions about what learning goals might be the most important for the student, and about what a student is able to accomplish in a set amount of time (say, a school term or two).

Planning for achievement: What works?

Once a teacher has determined the student's learning goals within a domain, the next step is to identify how best to support that student to achieve them. While teachers who are highly experienced in a domain and in teaching will likely draw on their expertise to plan for individual student achievement, many teachers may wish to seek the advice and support of other more experienced teachers to assist in their planning. Using evidence-based practices from research literature are also likely to support student learning, though these can be difficult to locate due to time pressure on teachers and a lack of access to research journals.

One way that the SWANs/ABLES program (University of Melbourne, 2017, 2018) supported teachers was to provide a list of curated evidence-based teaching strategies and advice that was targeted at each level of each learning progression for the nine foundational areas. These strategies and advice were jointly developed from research findings of impactful practices and the knowledge of expert teachers before being trialled with classroom teachers who had a range of experience and expertise in the domain areas as well as in teaching students with disability and/or additional learning needs. As a result, a teacher using the SWANs/ABLES program (University of Melbourne, 2017, 2018) to identify their student's current and next likely learning is also provided access to a range of targeted strategies and advice to support that student to reach their learning goals. While certainly not prescriptive, it can support teachers to not only recognise the good practices that they may already be implementing, as well as to suggest other new and different approaches. Teachers are encouraged to only apply strategies and advice that they perceive as relevant to their student, and to adapt strategies and advice as necessary to suit the individual needs or preferences of their students.

Understanding impact: What worked?

After a reasonable amount of time, such as a school term, a teacher should re-assess their student to determine the extent of learning against the goals that were set for that student. The same assessment procedures and tools used to determine the student's initial location on the learning progression should be used for subsequent assessments for reasons of reliability. These could be the same rubric or assessment tool, such as the SWANs/ABLES assessment tools (University of Melbourne, 2017, 2018), so that any demonstration of increasing capability by the student is recognised and acknowledged on the same scale that was used to determine their previous capability and to set their learning goals. Tracking student growth using the same metric allows for increased confidence in the result and fidelity in the teaching and learning cycle.

Once a student's learning growth has been assessed, it is important to consider the amount of growth as well as the possible reasons behind that amount of growth. Wherever possible, it can be helpful for teachers to have a clear representation of that growth, particularly when comparing growth across different domains. The SWANs/ABLES (University of Melbourne, 2017, 2018) individual reporting format uses a simple vertical arrow divided into sections to denote each level of a domain's learning progression, with thick black bars showing a student's estimated location on each assessment date. In this manner, teachers can quickly view growth between each time point in a domain. Another SWANs/ABLES (University of Melbourne, 2017, 2018) reporting format shows an individual's growth across multiple domains at each time of assessment, allowing for comparisons across domains, while other reporting formats allow students to be compared across a class or school, for the purpose of understanding general student abilities and to plan for different types of group learning.

When seeking to understand why a certain amount of growth has occurred, it is necessary to revisit the teaching strategies and advice used to inform the student's learning, as well as other factors such as resourcing, time, student absences, and other circumstances which may have positively or negatively affected learning. Teachers may wish to ask themselves, 'What evidence do I have that a particular intervention was successful – or not?'; 'Why did this student achieve so much in literacy, but relatively less in numeracy?'; 'What other aspects impacted their learning this term?'. Using these reflective teaching practices, supported by evidence, can help teachers to craft a better understanding of the student as a learner. They can also serve as the impetus for investigations, as appropriate, as to why a student may have achieved far more, or less, than expected in the time between assessments. Most importantly, they can help the teacher to craft increasingly effective teaching practices for each learner, so to support the success of every student.

Building on success

Teachers who incorporate high-quality learning progressions, such as those developed within SWANs/ABLES (University of Melbourne, 2017, 2018), into their practices can feel confident in providing an evidence-based, targeted learning experience for each student based on student capability, not age- or grade-level statements of what should be taught. The concept of the ZPD (Vygotsky, 1978) provides a strong basis for understanding student capability for the purpose of making good planning and teaching decisions for maximum growth. The use of evidence-based teaching strategies and advice to scaffold student achievement towards goals can mean that students make the most of their learning time, and meet high but reasonable expectations for their learning. Lastly, the implementation of reflective teaching practices to interrogate the reasons for student growth allows teachers to not only develop their own capacity as professionals, but to continue to provide the most effective teaching and learning for each student in their class. It is little wonder that organisations as large and respected as UNESCO see the implementation of learning progressions as central to the success of students worldwide; Australian students certainly deserve no less.

References

Griffin, P. (2014). *Assessment for teaching*. Cambridge University Press.

Heritage, M. (2008). *Learning progressions: Supporting instruction and formative assessment*. http://www.ccsso.org/documents/2008/learning_progressions_supporting_2008.pdf

Marope, M., Griffin, P., & Gallagher, C. (2019). *Future competences and the future of curriculum: A global reference for curricula transformation*. http://www.ibe.unesco.org/sites/default/files/resources/future_competences_and_the_future_of_curriculum.pdf?fbclid=IwAR1Y3Ac0dNgpHqIe02-OwYQ-b-eAZfwEHJ3PwnbUZa8TXIa-PZQpTrMjnhU

Stevens, S., Shin, N., Delgado, C., Krajcik, J., & Pellegrino, J. (2002). *Using learning progressions to inform curriculum, instruction and assessment design*. [Paper presentation]. NARST Annual International Conference, New Orleans, LA, United States. http://hi-ce.org/presentations/documents/Shawn_etal_NARST_07.pdf

University of Melbourne. (2017). *Assessment and teaching of foundational skills for diverse learners: The Students with Additional Needs program*. http://education.unimelb.edu.au/arc/projects/current/assessment_and_teaching_of_foundational_skills_for_diverse_learners_the_students_with_additional_needs_program#about

University of Melbourne. (2018). *Abilities Based Learning and Education Support (ABLES) research*. https://education.unimelb.edu.au/research/projects/abilities-based-learning-and-education-support-ables-research

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds.). Harvard University Press.

White, E. H. (2019). *Digital literacy for students with disability: Recognising ability and supporting learning through assessment, evidence, and expertise*. [Doctoral dissertation, The University of Melbourne, Melbourne, Australia]. http://hdl.handle.net/11343/224996

Wiliam, D. (2007). Content then process: Teacher learning communities in the service of formative assessment. In D. B. Reeves (Ed.), *Ahead of the curve: The power of assessment to transform teaching and learning*. Solution Tree Press.

Woods, K. (2010). *The design and validation of measures of communication and literacy to support the instruction of students with learning disabilities*. [Doctoral thesis, University of Melbourne, Melbourne, Australia].

Evidencing creativity and curiosity in IB schools

Dr Sarah Richardson
Australian Council for Educational Research

Dr Sladana Krstic
Australian Council for Educational Research

https://doi.org/10.37517/978-1-74286-638-3_7

Dr Sarah Richardson is a passionate researcher, educator and writer. She has been working in the education sector for 28 years as a teacher, lecturer, researcher and research director. Sarah is the Research Director of the Surveys and International Assessments team in the Australian Council for Educational Research (ACER). She returned to Australia in April 2021 having been Deputy CEO and Research Director of ACER UK, and previously Research Director of ACER India. Sarah has a PhD in education and politics from the University of Melbourne, a Master's Degree in International Relations from the University of Amsterdam and a Bachelor's Degree in Human Geography from the University of Liverpool.

Dr Sladana Krstic leads the psychometrics and research team for ACER UK. She has worked on a variety of projects, including the Scottish National Standardised Assessments; the British Council's English Impact project; the International Baccalaureate programmes; and the Royal Society of Chemistry evaluation programme. Sladana has a Doctorate in Education and is an expert researcher in quantitative and qualitative methodology, data analysis, sampling designs, and questionnaire development. She is currently engaged in a ground-breaking research project for the International Baccalaureate Organisation and the Jacobs Foundation on curiosity, creativity and learning.

Abstract

There is growing recognition of the importance of learners gaining transversal or 21st-century attributes in order to thrive in the contemporary world. This poses a number of challenges for educators. First, to what extent are transversal attributes innate, or do they include a combination of traits and skills? Second, what can teachers do to help nurture these attributes in learners? Third, how can the existence or strengthening of attributes be recognised? In this paper, we draw on work that we are doing for the International Baccalaureate Organisation to define conceptual frameworks for creativity and curiosity. Our goal is to enable learners to evidence achievement in each of these attributes, and for teachers to be able to recognise it. The frameworks draw on extensive scholarly literature to define the core components of both curiosity and creativity and the skills that are inherent in each one. This is the first step towards the development of a 'transversal résumé' that will allow learner achievement to be recorded. This includes a focus on the extent to which schools and teachers can provide learners with opportunities to gain, practice or enhance the skills that contribute to transversal attributes and a scaffold that enables learners to reflect on the extent to which they have evidenced creativity or curiosity in sustained pieces of work. The paper raises some key questions that have arisen as we have grappled with both conceptual and practical issues in this project. These provide valuable insights into the nurturing of transversal attributes, and the implications for educational professionals.

Introduction

There is a growing focus in education systems around the world about the importance of transversal attributes. Educational professionals are increasingly acknowledging that for learners to thrive in their future lives, they need to be equipped with a good balance of academic competencies and transversal attributes. How this can be achieved is, however, much less clear cut. There is no consensus about how to refer to these attributes – with alternatives being '21st century', 'holistic', 'generic', 'soft' and 'multi-disciplinary'. There is no consensus on whether these are even attributes at all, with some suggesting that they are skills.

In this context, the Australian Council for Educational Research (ACER) was approached by the International Baccalaureate Organisation (IBO) to develop a way to evidence two transversal attributes across its educational programmes. Generously funded by the Jacob's Foundation, ACER has been working closely with the International Baccalaureate (IB) since August 2020 to develop frameworks for curiosity and creativity. Our approach builds on previous work to develop frameworks in other transversal areas such as collaboration (Scoular et al., 2020) and critical thinking (Heard et al., 2020).

In this paper, we provide an overview of the process followed, an insight into the frameworks developed and a recognition of some of the significant debates that have made this work both intellectually and practically challenging.

Background and context

This work was informed by a number of key parameters. First, it needed to be usable across all IB programmes: the Primary Years Programme (PYP); the Middle Years Programme (MYP), the Diploma Programme (DP) and the Career-Related Programme (CP). This meant it had to be applicable for learners aged 3 to 19 years. Second, it needed to be relevant for all schools implementing IB programmes, wherever they are in the world. Third, it should enable schools to evidence learner achievement and progress without assigning grades or using formal assessment tools.

IB programmes have a foundational core philosophy, yet each retains a distinct focus and purpose. Broadly, IB programmes focus on holistic education that:

> … encourage both personal development and academic achievement challenging students to think critically, to ask the right questions and think across disciplines. An IB education also fosters diversity, curiosity and a healthy appetite for learning (International Baccalaureate Organisation, n.d.).

Schools that wish to provide IB programmes first need to be authorised to do so. Some schools offer just one programme and others more. As of May 2021, there are 5500 IB World Schools in 159 countries (International Baccalaureate Organisation, 2021). While diverse in nature, each programme is supported by the IB Learner Profile, which identifies a set of attributes that IB learners strive to have. These include being reflective, open-minded, thinkers and communicators. (International Baccalaureate Organisation, 2013). Another core element of all IB programmes is the focus on inquiry-based learning. While neither creativity nor curiosity are explicitly mentioned as attributes in the Learner Profile, they are key elements of much of the IB's philosophy.

While many education systems and programmes reference transversal attributes, it is common for this to be superficial, with little follow-through to the level of the curriculum, to school leadership, to teacher professional learning or to pedagogy. In this project with the IBO, ACER is focusing on developing resources that enable creativity and curiosity to be supported, that guide teachers in helping learners gain the skills that underlie these attributes and that enable achievement to be recorded.

In parallel, the Oxford University Centre for Educational Assessment is undertaking a project to gather samples of promising classroom practices where learners are provided with opportunities to gain or practice either creativity or curiosity. The two projects have worked closely together, with efforts to optimise synergies between them. Inevitably, both projects have been impacted by the Covid-19 pandemic and required adaptations to methodologies.

Methodology

Before a transversal attribute can be evidenced, it first needs to be defined. It is common practice to start with the development of a framework to propose a working definition and then to break down the overall construct into what are commonly referred to as 'strands' – overarching conceptual categories for framing the skills and knowledge associated with the construct. These are amenable to instruction, provide evidence of the construct, and can potentially be assessed. The strands are subsequently broken down into 'aspects' which are specific content categories within a strand.

As with many transversal domains, there is extensive scholarly literature about both creativity and curiosity. However, it can be extremely challenging to take something that may be esoteric and theoretical, or based on psychological and neuroscientific processes that are intangible, and define it in a way that is suitable for an educational context. Accordingly, this project started with a global three-day virtual symposium that brought together scholars with expertise on either curiosity or creativity, as well as educational professionals from IB schools around the world.

Through a series of facilitated discussions and reflections, the symposium was able to set a clear direction for the subsequent stages of the project. This included identifying three tensions that are often absent from the literature on transversal attributes in education. First, unless schools provide a suitably enabling environment for the practice of transversal attributes, it is very difficult for learners themselves to demonstrate the development of these. Hence, part of the evidencing of transversal attributes is for schools and teachers to demonstrate that they provide optimal conditions for transversal attributes to be nurtured.

Second, teachers play a key role in assisting learners to enhance transversal attributes and their corresponding skills, including through modelling these themselves. Given that emotions such as disappointment and frustration are inevitable parts of many transversal attributes, this means that teachers need to be provided with an environment in which learners are allowed to fail. Third, efforts to evaluate, measure or assess transversal attributes carry the inherent risk of reducing what is recognised as their expression, hence contradicting the nurturing of the attributes. In this project, we have steered clear of the terms 'measurement' or 'assessment', instead referring to these as 'evidencing', partly as an acknowledgement of this third tension.

The symposium was the starting point for work on this project. It was followed by three key steps, the third of which remains in progress at the time of writing. First, two teams of ACER researchers carried out an in-depth literature review of scholarly work on either creativity or curiosity. These literature reviews drew on hundreds of journal articles and provided a foundation for the development of frameworks. Second, the literature reviews were handed over to two teams of ACER framework experts, all of whom had previously worked on the development of frameworks for transversal domains. They drew on the literature to arrive at working definitions of each construct and an appropriate set of strands and aspects.

In the third step, ACER researchers commenced an extended period of consultation. This has involved scholarly experts from the initial symposium, IB curriculum and subject managers from each IB programme and senior IB managers. The next stage – somewhat delayed and requiring methodological workarounds due to the impact of the COVID-19 pandemic on schools – is to consult with teachers and programme coordinators in IB schools in all parts of the world. This paper provides an insight into the frameworks as they exist in June 2021. It is important to note that these are in draft form and may change substantially when consultations have concluded.

Draft frameworks

For each construct of curiosity and creativity, the first element of the framework is the definition. This ensures both a sound evidentiary basis as well as a consideration of factors that are amenable to teaching and learning. These definitions are then further elaborated using strands and aspects.

For creativity, the working definition focuses the attention on working with things, ideas, or people – or any combination of the three – in a purposeful and directed way, taking account of real-world constraints. The outcomes or 'products' of creativity are defined as having the key characteristics of novelty and usefulness. Beyond generating these, our definition posits that creativity involves openness to the exploration of ideas, including through the examination of these from different perspectives, and the synthesis of different ideas.

> Creativity is a material, mental and/or social process that leads to the production of novel and useful ideas, approaches and solutions. It involves the exploration, generation and evaluation of both problems and ideas, made possible through divergent, experimental and convergent thinking. (Scoular & Ramalingam, 2021).

The strands and aspects of creativity are defined as: problem finding (comprising discovery-oriented behaviour and formulating a problem); generating ideas (comprising fluency, flexibility and experimentation); and quality of ideas (comprising originality, fitness for purpose and elaboration).

For curiosity, the working definition focuses the attention on curiosity as an attribute that fosters deep and lifelong learning. It refers to a meaningful gap in which one addresses a substantive task that facilitates richer, deeper or broader conceptual understanding. Curiosity is regarded as an intrinsic part of a process of self-motivated development of knowledge and understanding. This includes many elements that are already emphasised in IB programmes, including inquiry skills, critical thinking, open-mindedness, risk taking, self-efficacy and reflection. Curiosity provides the incentive and reward for seeking new knowledge and understanding, requiring other skills to ensure that this learning is rich, sustained and productive.

> Curiosity involves the recognition of a meaningful gap in one's knowledge or understanding, the desire to fill that gap and the motivation and intrinsic satisfaction of doing so (Heard & Anderson, 2021).

The strands and aspects of curiosity are defined as: focusing curiosity (comprising engaging with and exploring conceptual conflicts, enhancing motivation and refining questions of value) and resolving knowledge gaps (comprising exploring answers and thinking critically, sustaining effort and evaluating learning).

Evidencing and enabling environments

Frameworks developed for cognitive skills, such as mathematics or reading, would normally specify different levels of competence or achievement, with the expectation that these are subsequently validated by the collection of assessment data. For transversal attributes, however, very little is known about how these are developed and this makes the development of a progression challenging. During the commencement of this project with the global symposium, participants cautioned against a reductionist approach that would define a construct narrowly enough to allow it to be assessed, hence squeezing the inherent freedom and space to explore out of these elements.

At the same time, the IBO is keen to be able to develop something that enables learners' transversal attributes to be summarised, albeit avoiding 'grading' them.

In this context, and after much debate and consideration, our approach has been to consider three elements: 1) the environment in which learning takes place (for schools); 2) opportunities to gain, practice or enhance the skills that are inherent within transversal attributes (for teachers); and 3) a scaffold that enables learners to reflect on the extent to which they have evidenced creativity or curiosity in sustained pieces of work (for learners). The underlying philosophy to our approach is a recognition that schools and teachers have a responsibility to provide the conditions in which learners have the time, space, permission and learning opportunities to be able to be creative and to be curious. To achieve these three elements, we are planning to develop the following.

For schools: A framework for reflection upon the environment for learning

A framework will help schools to reflect on the extent to which they provide enabling environments for creativity and curiosity. For elements such as 'task opportunities' and 'access to resources', we plan to provide schools with a number of 'temperatures' or 'levels' for their practices to be evaluated against. For example, schools that provide comprehensive or minimal opportunities for learners to be curious could recognise themselves as those in which:

> **Comprehensive** – Learners are given open tasks with clear guidelines that focus on promoting curiosity-driven learning; sufficient time is allowed for investigations and reflections; and curiosity not compromised by other task purposes.
> **Minimal** – Any opportunities to demonstrate curiosity are largely prescribed or predictable, or may be missing entirely (Heard & Anderson, 2021).

For teachers: Suggested approaches to support learners in gaining, practicing or enhancing the skills within transversal attributes

We plan to develop suggestions for teachers around how they can utilise classroom activities to support learners to gain, strengthen or practice the skills that facilitate creativity or curiosity. For example, in order to support learners to practice refining questions of value (an important aspect within curiosity) teachers may wish to focus on helping them to develop 'high-value questions that are likely to challenge understanding and lead to significant new insights' (Heard & Anderson, 2021). This may include elements such as the ability to distinguish between questions that address confusion or gaps in understanding, and those that might be secondary or consequential questions.

For learners: A structure that facilitates reflection

Learners will be facilitated to reflect (with support from teachers where required) on the extent to which their curiosity or creativity is embodied in sustained pieces of work using a structure. We consider that the best opportunities for this to occur would be in the extended culminating project in each programme, including the 'Exhibition' in PYP and 'Extended essay' in DP. For example, a learner may reflect on the extent to which they have engaged with, and explored, conceptual conflicts. This could involve identifying the extent to which they have practiced in-depth, open and imaginative exploration of stimuli; identified gaps, inconsistencies or contradictions from a range of perspectives and possibilities; and made connections with a broad range of ideas.

Conclusion

While this project remains in progress, with consultations with schools still to be undertaken, it has already raised some key questions around transversal attributes and their incorporation into educational settings. First, it has identified the importance of schools providing an enabling environment in which learners – and teachers – are free to be creative and curious. Second, it has identified the need to consider the ways in which teachers can help learners acquire, practice and strengthen the skills that contribute to creativity and curiosity. Third, it has highlighted the challenges involved in evidencing creativity and curiosity, and the desirability of engaging learners in reflecting on their own achievement.

While the project focuses on two specific transversal attributes, we feel that these issues are equally important considerations for other transversal attributes. Education sectors are becoming increasingly aware of the need to prepare learners for their future lives, ones that will demand a wide variety of attributes, skills and knowledge. Unpacking, defining and operationalising transversal attributes requires critical evaluation of the implications for educational institutions and professionals. This project makes an important contribution to that debate.

The authors would like to thank the Jacobs Foundation and International Baccalaureate Organisation for funding this project. We would also like to recognise the contributions of our current and former ACER colleagues Prue Anderson, Anit Cherian, Laura Good, Jonathan Heard, Brendan McGinley, Abigail Middel, Dara Ramalingam, Claire Scoular, Shani Sneidze-Gregory, Jennifer Star and Anaghaa Wagh.

References

Heard, J. & Anderson, P. (2021). *DRAFT curiosity combined preamble, framework and progression* [Manuscript in preparation]. Australian Council for Educational Research.

Heard J., Scoular, C., Duckworth, D., Ramalingam, D., & Teo, I. (2020). *Critical thinking: Skill development framework.* Australian Council for Educational Research. https://research.acer.edu.au/ar_misc/41

International Baccalaureate Organisation (nd). *Programmes*. https://www.ibo.org/programmes/

International Baccalaureate Organisation (2013). *IB learner profile.* https://www.ibo.org/contentassets/fd82f70643ef4086b7d3f292cc214962/learner-profile-en.pdf

International Baccalaureate Organisation (2021). *Facts and figures*. https://www.ibo.org/about-the-ib/facts-and-figures/

Scoular, C. & Ramalingam, D. (2021). *DRAFT Creativity combined preamble, framework and progression* [Manuscript in preparation]. Australian Council for Educational Research.

Scoular, C., Duckworth, D., Heard, J., & Ramalingam, D. (2020). *Collaboration: Skill development framework*. Australian Council for Educational Research. https://research/acer/edu.au/ar_misc/42

Exploring excellence in Indigenous education in Queensland secondary schools

Dr Marnee Shay
The University of Queensland

Dr Jodie Miller
The University of Queensland

Dr Suraiya Hameed
The University of Queensland

https://doi.org/10.37517/978-1-74286-638-3_8

Dr Marnee Shay is an Aboriginal researcher, with maternal connections to Wagiman country and strong ties to Aboriginal communities in south-east Queensland where she was born and raised. Having a diverse professional background, she has worked in the community sector as a youth worker and as a classroom teacher in secondary schooling and tertiary contexts. Dr Shay is nationally recognised for her leadership in education research. She is an interdisciplinary researcher who has been awarded multiple externally funded projects across the fields of Indigenous education, youth studies and policy and practice.

Dr Jodie Miller is a non-Indigenous education researcher working as a Senior Lecturer at the University of Queensland. Prior to working in universities Jodie was a primary classroom teacher. Her research area focuses on improving the teaching and learning of mathematics, particularly for students who are most at risk of marginalisation from the school mathematics curriculum. In addition to this, her recent research collaborations focus on examining excellence in Indigenous education.

Dr Suraiya Hameed is a Lecturer in Educational Leadership at the University of Queensland. She has Indigenous Malay ethnic roots, on her maternal side within the Singaporean context. She has been in education for 30 years and was previously a school leader before transitioning into a lecturing and research position at UQ. Her research interests are in exploring complex educational issues – from shifts in cultural practices and racial, ethnic, and linguistic diversity to the implications of these shifts for human or social development and education. Her current research focus is within the area of leadership, global citizenship & inclusive education specifically in the area of Indigenous education.

Abstract

In the national and international landscape, there is very limited exploration of cultural constructs of excellence, in particular, in Indigenous contexts. This pilot study aimed to centre the voices of Indigenous people in conceptualising excellence in Indigenous education, as well as to share understandings between Indigenous and non-Indigenous practitioners. Qualitative data collection methods were used including collaborative yarning, storying, and semi-structured interviews. Data were analysed using cross-case analysis to examine the views of educators across three school sites. Indigenous participants highlighted the importance of nurturing culture and identity; building up young people; and, building a culture of inclusivity and belonging. Supportive leadership was also identified as an enabler for enacting excellence in schools. A direct outcome of this project was a whole-school policy that builds on a strengths perspective and forefronts the embedding of Indigenous knowledges and perspectives, supporting the wellbeing of Indigenous students, affirming the identities of Indigenous students and having specific strategies to engage with local Indigenous communities.

Introduction

The terms ‘Indigenous education’ and ‘excellence’ in Indigenous education are so conceptually distanced, that when we commenced this project, a good search of the two terms yielded little information. The distance between these concepts in scholarly and policy domains demonstrates the depth and pervasiveness of deficit discourses in Indigenous education in Australia. The term excellence has emerged from Indigenous communities within Australia as a way of recognising and acknowledging strengths, knowledges and values of Indigenous peoples and communities in the face of the persistence of deficit narratives in social, political and educational discourses. In small pockets of social media or in the naming of various educational programs, for example, the term Indigenous excellence is operationalised by Indigenous peoples in various contexts. We have attempted to understand what Indigenous education excellence is and the ways in which it is conceptualised by centring Indigenous voices but also including the voices in education who are responsible for enacting Indigenous education policies: principals, teachers and education workers. It is important to distinguish that we are not defining Indigenous excellence; we are exploring the concept of excellence in the many elements that make up Indigenous education practices.

A synthesis of the literature that focuses on Indigenous learners in schools revealed six key themes where research is focused: these predominantly cover issues on identity, cultural competence of educators, engaging with Indigenous families and communities, presence of Indigenous cultures in schools, employment of Indigenous peoples in schools, and leadership (Shay & Heck, 2016). There is also increasingly a focus on the embedding of Indigenous knowledges and perspectives into curriculum; advocating the need for it, critical discussion about the contestability of Australian histories and the ability to embed across all curriculum areas, and these have all emerged in national discussions on the issue (Phillips & Lampert, 2012; Sarra, 2011; Smith et al., 2019; Sarra & Shay, 2019).

The concept of excellence in education is relatively undertheorised. Walker (1996) highlights complexities in understanding the term in relation to the subjectivities in how the term excellence is defined, understood and applied. The use of the term excellence broadly in education has often lacked a holistic view of what it can include and emphasises academic merit over other educative endeavours and outcomes. The OECD states clearly that the promotion of ‘excellence, equity and inclusion are key aims for education’ (Schleicher, 2014, p. 11) and is a priority for the global community. However, measurement of excellence undertaken through PISA (Programme for International Student Assessment) is through testing students on their science, maths, reading, collaborative problem-solving skills and financial literacy capability (OECD, 2018). In the national and international landscape, there is very limited exploration of cultural constructs of excellence, in particular, in Indigenous contexts.

Background

This paper shares key findings from a pilot study, ‘Doing things right way: Dimensions of excellence in Indigenous education in Queensland secondary schools’. This project was led by Aboriginal researcher Shay, Miller (non-Indigenous Australian researcher, and Hameed (an Indigenous researcher from Singapore). We explored what excellence in Indigenous education is or could be. This study aimed to centre the voices of Indigenous people in conceptualising excellence in Indigenous education, as well as to share understandings between Indigenous and non-Indigenous practitioners.

The following research questions were used to study participants' conceptualization and enactment of excellence at school level.

1. How is excellence in Indigenous Education defined by Indigenous educators and community leaders?
2. What are some examples Indigenous educators and community leaders identify as excellence in Indigenous Education?
3. What do Indigenous educators and community leaders indicate as ways leaders can support the enactment of excellence in Indigenous education?

The pilot study explored dimensions of excellence in Indigenous education and demonstrated the need to examine the topic of excellence in Indigenous education in a wide variety of contexts where Indigenous education is being practised (Shay & Miller, 2019). The key findings demonstrated that the two terms 'excellence' and 'Indigenous education' have been so conceptually distanced that both Indigenous and non-Indigenous participants often reverted back to discourses of gaps and deficits even when being explicitly asked what excellence in Indigenous education is or could be.

Methodology

This was a qualitative study, with a collective case study methodology underpinning the design of the project, allowing for an in-depth understanding of the phenomenon and multiple examples that are contextualized and localized (Punch, 2013). Cross-case analysis was conducted to sieve out similarities and differences in the way the schools are interpreting and realising Indigenous education excellence. Emergent meta-level conceptual themes on Indigenous education excellence were discussed: enablers and constraints; and the relevance of distinctions between Indigenous excellence and academic achievement were also addressed.

Qualitative data from various means were collected. A number of methods were used in storying the case studies. The use of yarning, an Aboriginal conversational style of sharing knowledges and experiences (Bessarab & Nga'ndu, 2010) were utilised to develop ways of identifying and understanding excellence in Indigenous education with both Indigenous and non-Indigenous participants. Yarning in documenting the case study from multiple perspectives included semi-formal questions about the practice, collaboration with Indigenous peoples in the process, the resourcing required, and enablers. Data were recorded using a methodology developed by Shay (2019), known as collaborative yarning methodology. Recording of yarns via a storyboard enables a more culturally relevant and ethical way of incorporating diverse Indigenous voices.

The theoretical lens used is based on Rigney's Indigenist principles of political integrity (2001), resistance as the emancipatory imperative and privileging of Indigenous voices in the conceptual framing of Indigenous research. While the data include the perspectives of non-Indigenous researchers (who make up the majority of the education workforce), the research design ensured there were mechanisms to aim for at least half of all participants to be Indigenous, ensuring that Indigenous peoples' perspectives, stories, experiences and aspirations were central in conceptualizing what Indigenous education excellence is or could be. Specific analysis to foreground the voices of Indigenous perspectives across all research questions was informed by Rigney's Indigenist principles (2019).

Discussion

Indigenous participants in this Study described their conceptualisations of excellence in Indigenous education within three key areas: *nurturing culture and identity; building up young people; and, building a culture of inclusivity and belonging*. Other themes that surfaced included an emphasis on relationships, supporting young people to achieve goals, and walking together.

It was clear in the data that exploring Indigenous education through a lens of excellence was a foreign notion to some participants. Our analysis suggests that this was due to the prevailing deficit discourses that exist in policy, scholarship and practice and these discourses have also been absorbed by Indigenous peoples in education settings (though to a lesser extent). As this research is original in concept, these initial findings suggest that growing this data set will assist in producing a broader understanding and will assist in theorizing what Indigenous education excellence is or can be in Australia. The implications for this Study are far reaching, as educational parity for First Nations students globally remains a critical social justice issue.

Another key finding within the study is the important role that leadership plays in the enactment of excellence in schools. The findings from the literature review and also the empirical research conducted showed that there is a need for school leaders who are cognisant of the importance of providing a conducive environment that respects and values Indigenous knowledges, have high expectations for Indigenous students' achievements and most importantly encourage meaningful and culturally responsive pedagogical practices that assist in the building of a strong culture and enhance learning and involvement for Indigenous students (Hameed, Shay & Miller, 2021).

Many Indigenous educators also discussed the importance of having supportive school leaders who provided them with opportunities to lead in their school communities. What became apparent through our yarning sessions was the importance of leaders' embedding Indigenous education, culture and histories into the life of the school. Leaders needed to be serious about: 1) embedding this into their school policy documents; 2) moving beyond singular occurrences in school life; and, 3) creating a sustainable ongoing commitment to acknowledging and supporting Indigenous young people in their schools (Hameed, Shay & Miller, 2021).

Developing a school-wide policy framework

To support educational leaders in developing a school-wide policy framework for implementing Indigenous education, Shay & Miller (2021) published a chapter in the book *Building better schools with evidence-based policy* (Allen et al., 2021). The school-based policy is informed by strengths approaches and data developed through this Study. This chapter intentionally employs language that shifts away from notions of 'closing the gap' to endorsing existing research connected to key findings from our Study about how excellence in Indigenous education was conceptualised. Key features of the policy include intentions of embedding Indigenous knowledges and perspectives, supporting the wellbeing of Indigenous students, affirming the identities of Indigenous students and having specific strategies to engage with local Indigenous communities.

References

Allen, K.-A., Reupert, A., & Oades, L. (Eds.). (2021). *Building better schools with evidence-based policy: Adaptable policy for teachers and school leaders*. Routledge. https://doi.org/10.4324/9781003025955

Bessarab, D., & Ng'andu, B. (2010). Yarning about yarning as a legitimate method in Indigenous research. *International Journal of Critical Indigenous Studies*, *3*(1), 37–50.

Hameed, S., Shay, M., & Miller, J. (2021, forthcoming). Deadly leadership in the pursuit of Indigenous education leaders. In *Future alternatives for educational leadership: Diversity, inclusion, equity, and democracy.* Routledge.

Phillips, J., & Lampert, J. (2012). *Introductory Indigenous studies in education* (2nd ed.). Pearson.

Punch, K. (2013). *Introduction to social research: Quantitative and qualitative approaches*. Sage.

Rigney, L. (2001). A first perspective of Indigenous Australian participation in science: Framing Indigenous research towards Indigenous Australian intellectual sovereignty. *Kaurna Higher Education Journal*, *7*, 1–13.

Sarra, C. (2011). *Strong and smart – Towards a pedagogy for emancipation: Education for first peoples*. Routledge.

Sarra, G., & Shay, M. (2019). Indigenous education, critical perspectives to enhance learning practices. In *Encyclopedia of teacher education* (pp. 1–8). Springer. https://doi: 10.1007/978-981-13-1179-6_195-1

Schleicher, A. (2014). *Equity, excellence and inclusiveness in education: Policy lessons from around the world*. OECD.

Shay, M. (2019). Extending the yarning yarn: Collaborative yarning methodology for ethical Indigenist education research. *The Australian Journal of Indigenous Education*, 1–9. https://doi.org/10.1017/jie.2018.25

Shay, M., & Heck, D. (2015). Alternative education engaging Indigenous young people: Flexi schooling in Queensland. *The Australian Journal of Indigenous Education*, *44*(1), 34–47. https://doi: 10.1017/jie.2015.8

Shay, M. & Miller, J. (2019, December 1–5). *Doing things right way: Dimensions of excellence in Indigenous education in Queensland secondary schools* [Paper presentation]. Australian Association for Research in Education, Brisbane, QLD, Australia.

Shay, M. & Miller, J. (2021). Excellence in Indigenous education. In K.-A. Allen, A. Reupert & L. Oades (Eds.). *Building better schools with evidence-based policy: Adaptable policy for teachers and school leaders* (pp.46–54). Routledge.

Smith, L., Tuck, E., & Yang, K. W. (2019). *Indigenous and decolonizing studies in education: Mapping the long view* (Indigenous and decolonizing studies in education). Routledge.

OECD. (2018). *Programme for International Student Assessment*. www.oecd.org/pisa/aboutpisa

Walker, R. (1996). Excellence with decency. *Equity and excellence in education*, *29*(1), 97–101.

Using assessment data to improve equity: How teachers use insights from the Scottish National Standardised Assessments

Dr Sarah Richardson
Australian Council for Educational Research

Dr Sladana Krstic
Australian Council for Educational Research

https://doi.org/10.37517/978-1-74286-638-3_10

Dr Sarah Richardson is a passionate researcher, educator and writer. She has been working in the education sector for 28 years as a teacher, lecturer, researcher and research director. Sarah is the Research Director of the Surveys and International Assessments team in the Australian Council for Educational Research. She returned to Australia in April 2021 having been Deputy CEO and Research Director of ACER UK, and previously Research Director of ACER India. Sarah has a PhD in education and politics from the University of Melbourne, a Master's Degree in International Relations from the University of Amsterdam and a Bachelor's Degree in Human Geography from the University of Liverpool.

Dr Sladana Krstic leads the psychometrics and research team for ACER UK. She has worked on a variety of projects, including the Scottish National Standardised Assessments; the British Council's English Impact project; the International Baccalaureate programmes; and the Royal Society of Chemistry evaluation programme. Sladana has a Doctorate in Education and is an expert researcher in quantitative and qualitative methodology, data analysis, sampling designs, and questionnaire development. She is currently engaged in a ground-breaking research project for the International Baccalaureate Organisation and the Jacobs Foundation on curiosity, creativity and learning.

Abstract

Evidence-based decision-making is regarded as an important indicator of quality in schools around the world. Using data gathered from assessments, in conjunction with other insights, can help school leaders and teachers better meet the needs of learners. In schools that cater to disadvantaged learners, using data to design targeted interventions plays an important role in improving equity. In this paper we report on a study with five schools in Scotland. All schools had learner cohorts characterised by multiple layers of disadvantage. Informed by the theoretical underpinnings of sensemaking theory, we investigated how teachers and school leaders used data from the Scottish National Standardised Assessments (SNSA). Our findings suggest that teachers and leaders are adept at combining assessment data with other insights – including their own observations. All schools were active in using data to inform decision-making, both at the whole-school level and at the classroom level. They reported multiple uses of data, from validating their own instincts to targeting support to particular cohorts of learners. We suggest that the way in which SNSA is designed – explicitly providing data to teachers to help inform their professional judgement – is a factor in the positive approach to data usage among these schools.

Introduction

There is a growing focus on the importance of evidence-based approaches to inform educational decision-making. Large-scale assessment is a key component of this, with the intent that it provides educational professionals with empirical insights that they can use to inform strategies that improve learning.

Unfortunately, research shows that many teachers struggle to apply data to their teaching practice (Cowie & Cooper, 2017; Selwyn, 2016) and that school leaders are unsure of how to support good data practices at their schools (Young et al., 2018). Overall, this tends to lead to a 'sorry mixture of confusion, technical naivety and misleading advice' (Demie, 2010, p. 446), resulting in some studies suggesting that many schools do not use data-based decisions (Silva et al., 2020).

In this study, we investigated how teachers and school leaders in Scottish schools use data to inform strategies to support improved student learning. This research was undertaken in a context in which the Scottish National Standardised Assessments (SNSA) were designed with a specific focus on providing teachers with detailed information about student learning, and support in applying this information to their professional practice. Our focus was on schools that catered to some of the most disadvantaged learners in Scotland.

Background and context

The Scottish Government's *Curriculum for Excellence* has established a new vision for education for Scottish learners (Education Scotland, 2020). One of its strategies is to deliver 'excellence and equity for learners'. Developed in 2017 within this framework, the SNSAs are a unique approach to large scale assessment, designed specifically to provide 'diagnostic information to support teachers' professional judgement' (Scottish Government, 2017).

The SNSAs are online, adaptive assessments for learners in Grades 1, 4, 7 and 10 of all government primary and secondary schools in Scotland and focus on numeracy, reading and writing. Teachers receive immediate diagnostic reports on individual learners and groups of learners and the goal is that they will 'assess children's progress and plan next steps in learning' (Scottish Government, 2017). Teachers are able to choose when to conduct the assessments, how to conduct them and what to do with the data they receive.

Teachers are offered training on how to implement the SNSA and what to do with the data. Using these data, teachers can identify precisely what their students (both as individuals and as a cohort) know and can do at the time of assessment, with content-specific descriptions providing rich insights into areas of strength and weakness.

The context in which the SNSAs are taking place is one of significant disadvantage. Scotland has a weaker performing education system in comparison with many European countries. In PISA 2018 (Programme for International Student Assessment), for example, Scottish secondary students scored just above the OECD average for reading, a significant drop since 2000 and 2009 (Scottish Government, 2019a). At the primary level, Scottish primary school students' reading levels were below that of many other European countries for reading in 2006 (and lower than the Scottish scores for reading in 2001) (Mullis et al., 2007) and their mathematics levels were below the mean overall Trends in International Mathematics and Science Study (TIMSS) score in 2007 (Mullis et al., 2008).

Educational underachievement is partly explained by the fact that almost a quarter of all children in Scotland live in poverty (Scottish Government, 2019b). In this context, many students have multiple layers of disadvantage (Kintrea, 2018). This means that teachers have challenging terrain to navigate in meeting the needs of students in their classes. With this background in mind, the ability to receive detailed diagnostic information on their students should enable teachers and schools to provide individual students with tailored interventions that meet their specific needs.

The question, however, is whether teachers and school leaders have the skills to use data from the SNSA to target educational interventions that address education inequity. In undertaking our research, we were informed by the conceptual framework of sensemaking theory, which focuses on the ways that individuals try to transform sources of information into a comprehensible interpretation that drives action (Weick et al., 2005). The concept has been applied in a number of contexts, including education. The use of data by teachers involves the making of meaning and the development of interpretations that can guide their work. (Riehl et al., 2018; Snodgrass Rangel et al., 2019; Vanlommel & Schildkamp, 2019).

Methodology

Our approach was to identify two of the most disadvantaged regions of Scotland (Glasgow and Renfrewshire) and to contact the relevant local government authority to gain permission to undertake research in schools. A total of five schools agreed to participate. To provide context for the findings, a summary of salient characteristics from each school is provided in Table 1. All numbers and percentage estimates were provided by the schools themselves.

Table 1 School characteristics

School	Level	Number of learners	Learners from SIMD 1–2[1] (%)	Learners with EAL[2] (%)	Learners with ASN[3] (%)
A	Primary	530	50–60	20–30	40–50
B	Primary	336	50–60	10–20	20–30
C	Primary	268	40–50	0–10	30–40
D	Primary	450	10–20	0–10	20–30
E	Secondary	1345	10–20	0–10	20–30

1 The first or second vigintile in the Scottish Index of Multiple Deprivation (SIMD)
2 English as an additional language
3 additional support needs

Once agreement was reached, two researchers (at least one of whom was a former teacher) spent one whole day in each school. They conducted semi-structured interviews with between one and three teachers at each school, in addition to a total of nine principal teachers, five head-teachers and nine deputy head-teachers across all schools. The teachers had an average of 16 years of teaching experience. The semi-structured interviews were based on a protocol that was sufficiently consistent to collect similar insights from all schools, while being flexible enough to engage in deep and naturalistic conversations with participants and to respond to the directions that discussions took.

All interviews were audio-recorded and the recordings were transcribed and then coded. The coding approach was informed by grounded theory (Glaser, 1978). The aim was to generate a descriptive and explanatory theory within the conceptual frame of sensemaking theory that can be used to derive practical implications for educational policy and practice.

Findings

Findings from the data yielded a large number of important insights. Most importantly, they highlighted the large number of ways in which teachers and school leaders were drawing on data: to monitor and track progress; to identify gaps and validate teacher judgements; to plan interventions and tailor teaching; to set expectations; and, to communicate outcomes with stakeholders. School leaders and teachers further provided evidence of the triangulation of data – drawing on a mixture of assessment data and other insights to inform decision-making.

All schools were found to have an active culture of using data to identify learner achievement. Most schools held regular attainment meetings between leaders and teachers in addition to meetings between the senior management team to utilise data in school planning. Within-school activities were reinforced by a data culture outside of schools, with head teachers attending peer-review meetings with other head teachers in the same local authority each term. In these meetings, attainment data were compared to peer schools in the local area and were also gathered by local authorities to use for benchmarking purposes.

Teachers described using a wide range of assessment types to facilitate triangulation including: classroom-based assessments and direct observations, formative and summative assessments, standardised and diagnostic assessments, such as SNSA, Progress in Reading Assessment (PiRA), teacher and parent questionnaires, and previous teacher reports and observations. Teachers also referred to the use of learner-centric sources of information aimed at capturing the learner's voice and experience, which one termed 'learning conversations with children'. As one teacher commented, 'data is brilliant but it's not good unless you have that voice that goes along with it'.

In terms of data used, all teachers interviewed expressed caution about the risk of over-burdening learners with assessment. Nevertheless, teachers also highlighted the value of accurate assessment data. One reported approach was to emphasise to learners that the assessment was for the teacher as well. Teachers reported that this encouraged learners to be honest and open in sharing the difficulties they were having. Teachers also reported explaining to their classes:

> We've done this learning from August to September, I want to make sure that you're all really confident so that I know where to go when we are moving on … you're doing this to inform me of where we go next – do we need to revise – … or can we power through?

Interestingly, moderation within and between schools, as well as peer-to-peer observations were also utilised as methods to 'draw a comprehensive picture' of the learner and ensure the quality of the conclusions in terms of the learner's learning. One school also used this as an opportunity to encourage teachers to participate in professional enquiry, including professional reading and research. While interviewees agreed they were 'data rich' in their schools, with multiple sources of data available, they were also cognisant of the impact of over-assessing learners and collecting data for its own sake. As one teacher commented: 'More [data] is not necessarily better data'.

While data from formal assessments were used in all schools, many teachers reported using different assessment methods to supplement formal data, such as a wide range of formative assessment methods that enabled them to adjust lessons as they were progressing, and to plan for subsequent ones. As one teacher reported:

> I use formative assessment all the time. I just pick children [to answer questions], and that gives me a good idea who's picking up the lesson, who's understanding what I am saying, and just through the talking and listening I can get that … [I ask for] thumbs up, thumbs down to show understanding of what I'm saying.

In addition to the variety of data used, teachers also reported using data for a number of different purposes. Interviewees noted using data for planning whole-school, year group, and class level activities. As a teacher explained:

> What we were looking to do was to identify those children who were just below average, to see if we could push them on, so that we could target our support around about that group.

In informing the design of interventions, interviewees reported using a range of sources of data, such as about learners' home lives, and combining these with assessment data, records of previous interventions and records from external agencies in order to target support holistically. As a head teacher noted, they try to identify:

> Who's not on track, why they're not on track, is it something to do with something that's happening with the families? Is it social work being involved? We look at all the interventions surrounding that one child, what's impacting on their success and their achievement and then we see what we can put in place for them.

In the same ways that schools reported using data to help learners who needed greater support, some also reported using this approach to group high-achieving learners. As one interviewee described:

> … we've got a challenge group set up because they were, you know, heads above the other wee ones with regard to understanding and their comprehension, so there's a wee challenge group that we've set up for that.

In addition to using data for planning purposes, teachers reported using assessment results to tailor their teaching approaches. Interviewees also reported using assessment data to monitor learners' performance and progression, and to identify any gaps that needed to be addressed with targeted interventions. These quotes reflect some of the uses of data:

> The teacher can use [data] to track and ensure that a child is still making positive progress and ensure there are appropriate targets being set for the children.

> We look at the attainment across the school, and we look at any gaps that there are. And try to come up with, as a team, why we think there are these gaps and what we are going to do about it.

Finally, and perhaps most importantly, teachers reported using assessment data to validate their own professional judgement, question anomalies and to work together with colleagues to examine data patterns and learn from one another. Two quotes illustrate these themes:

> As a teacher you always want to be assured that what you're doing is the right thing, and what you have thought is the right thing. I think it certainly does give peace of mind that our standards and expectations, the criteria for a certain level, is accurate.

> I mean obviously teacher judgement is - trumps everything - but we do find that all the data informs it really, really quite well.

Conclusion

This study drew on insights from teachers and school leaders about the ways in which empirical data from the SNSA, in conjunction with other information, supported their professional practice. In the schools involved in this study, data were used for multiple purposes and at multiple junctures throughout the school year. Ultimately, it is through the process of triangulation: using multiple sources of data, in conjunction with observations and knowledge of learners, that the most accurate judgements are achieved. The study indicates that having a holistic approach, in which assessment data is used as one indicator of learner ability but is not the sole determinant of it, is the most meaningful way to inform interventions and classroom practices.

In Scotland, the SNSA is a key assessment helping teachers understand the progress of each child, which informs planning for everyone, from teachers to the senior management team at each school. The supportive nature of SNSA – in which data are explicitly provided to teachers to help inform their professional judgement - is likely to be a factor in the positive approach to data usage. This is in contrast to many large-scale assessment programmes where data are used for accountability purposes, taking the element of judgement away from teachers and the profession (Lingard et al., 2017).

In all schools in this study, the use of empirical insights was not the responsibility of individual teachers in isolation, but a whole-of-school responsibility. This helps to foster teacher capacity and beliefs about data (Datnow & Hubbard, 2016) and promotes collective efficacy, perceived control and motivation (Prenger & Schildkamp, 2018). In a supportive environment, teachers are confident in making inferences that start from the question 'what do the data tell me?' (Coburn et al., 2009). In this way, making sense of data contributes to continuous improvement strategies to inform teaching and learning, something that can be considered a 'strategic sensemaking endeavour' (Park et al., 2012, p. 667).

References

Coburn, C., Honig, M. & Stein, M. (2009). What is the evidence on districts' use of evidence? In J. D. Bransford, D. J. Stipek, N. H. Vye, L. M. Gomez, & D. Lam (Eds.) *The role of research in educational improvement* (pp. 67–86). Harvard Educational Press.

Cowie, B., & Cooper, B. (2017). Exploring the challenge of developing student teacher data literacy. *Assessment in Education: Principles, Policy & Practice*, *24*(2): 147–163. https://doi.org/10.1080/0969594X.2016.1225668

Datnow, A. & Hubbard, L. (2016). Teacher capacity for and beliefs about data-driven decision-making: A literature review of international research. *Journal of Educational Change*, *17*:7–28. https://doi.org/10.1007/s10833-015-9264-2

Demie, F. (2010). Using value-added data for school self-evaluation: A case study of practice in inner-city schools. *School Leadership and Management*, *23*(4), 445–467. https://doi.org/10.1080/1363243032000150971

Education Scotland (2020). *What is Curriculum for Excellence?* https://education.gov.scot/education-scotland/scottish-education-system/policy-for-scottish-education/policy-drivers/cfe-building-from-the-statement-appendix-incl-btc1-5/what-is-curriculum-for-excellence #

Glaser, B. (1978). *Theoretical sensitivity: Advances in the methodology of grounded theory*. Sociology Press.

Kintrea, K. (2018). *Education, disadvantage and place in Scottish secondary education*, CR&DALL Working Paper CR&DALL WP301/2018, CR&DALL, Glasgow (UK).

Lingard, B., Sellar, S., & Lewis, S. (2017). Accountabilities in schools and school systems. *Oxford research encyclopaedia of education*. https://oxfordre.com/education/view/10.1093/acrefore/9780190264093.001.0001/acrefore-9780190264093-e-74.

Mullis, I., Martin, M., Kennedy, A. and Foy, P. (2007). *PIRLS 2006 international report*. TIMSS & PIRLS International Study Centre.

Mullis, I., Martin, M. and Foy, P. (2008). *TIMSS 2007 international mathematics report*. TIMSS & PIRLS International Study Centre.

Park, V., Daly, A. & Wishard Guerra, A. (2012). Strategic framing: How leaders craft the meaning of data use for equity and learning. *Educational Policy*, *27*(4), 645–675. https://doi.org/10.1177/0895904811429295

Prenger, R., & Schildkamp, K. (2018). Data-based decision-making for teacher and student learning: A psychological perspective on the role of the teacher. *Educational Psychology*, *38*(6), 734–752. https://doi.org/10.1080/01443410.2018.1426834

Riehl, C., Earle, H., Nagarajan, P, Schwitzman, T. & Vernikoff, L. (2018). Following the path of greatest persistence: Sensemaking, data use and the everyday practice of teaching. In N. Barnes & H. Fives (Eds). *Cases of teachers data use*. Routledge.

Scottish Government (2019a). *Programme for International Student Assessment (PISA) 2018: Highlights from Scotland's results*. https://www.gov.scot/publications/programme-international-student-assessment-pisa-2018-highlights-scotlands-results/pages/5/

Scottish Government (2019b). *Poverty and income inequality in Scotland: 2015–2018.* https://www.gov.scot/publications/poverty-income-inequality-scotland-2015-18/pages/4/

Scottish Government (2017). *Scottish National Standardised Assessments*. https://standardisedassessment.gov.scot/

Selwyn, N. (2016). 'There's so much data': Exploring the realities of data-based school governance. *European Educational Research Journal*, *15*(1), 54–68. https://doi.org/10.1177/1474904115602909

Snodgrass Rangel, V., Bell, E. & Monroy, C. (2019). Teachers' sensemaking and data use: Implementation in science classrooms. *Education and Urban Society*, *51*(4), 526–554. https://doi.org/10.1177/0013124517727053

Silva, M., Collier-Meek, M., Codding, R., Kleinert, W. & Feinberg, A. (2020). Data collection and analysis in response-to-intervention: A survey of school psychologists. *Contemporary School Psychology*. https://doi.org/10.1007/s40688-020-00280-2

Vanlommel, K. & Schildkamp, K. (2019). How do teachers make sense of data in the context of high-stakes decision-making? *American Educational Research Journal*, *56*(3), 792–821. https://doi.org/10.3102/0002831218803891

Weick, K., Sutcliffe, K. and Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, *16*(4), 409–421. https://doi.org/10.1287/orsc.1050.0133

Young, C., McNamara, G., Brown, M. and O-Hara, J. (2018). Adopting and adapting: School leaders in the age of data-informed decision-making. *Educational Assessment, Evaluation and Accountability*, *30*, 133–158. https://doi.org/10.1007/s11092-018-9278-4

Interpreting learning progress using assessment scores: What is there to gain?

Dr Nathan Zoanetti
Australian Council for Educational Research

https://doi.org/10.37517/978-1-74286-638-3_17

Dr Nathan Zoanetti joined ACER in 2018 as Research Director of the Psychometrics and Methodology Program and before that he was Principal Psychometrician at the Victorian Curriculum and Assessment Authority. He has previously held research positions at the Assessment Research Centre, University of Melbourne, including as Executive Officer of the NAP Literacy and Numeracy (NAPLAN) Measurement Advisory Group, and more recently as an Honorary Senior Fellow. He is currently a member of the OECD PISA 2021 Creative Thinking Expert Group and a member of England's Standards and Testing Agency Technical Advisory Group.

Abstract

Using assessment scores to quantify gains and growth trajectories for individuals and groups can provide a valuable lens on learning progress for all students. This paper summarises some commonly observed patterns of progress and illustrates these using data from ACER's Progressive Achievement Test (PAT) assessments. While growth trajectory measurement requires scores for the same individuals over at least three but preferably more occasions, scores from only two occasions are naturally more readily available. The difference between two successive scores is usually referred to as gain. Some common approaches and pitfalls when interpreting individual student gain data are illustrated. It is concluded that pairs of consecutive scores are best considered as part of a longer-term trajectory of learning progress, and that caveated gain information might at best play a peripheral role until additional scores are available for individuals. This review is part of a larger program of research to inform future reporting developments at ACER.

Introduction

Progress can be quantified using assessment scores as soon as two score points are available for the same individual. However, there are well-known technical shortcomings associated with quantifying progress based on only two scores (Willett, 1994; McCaffrey et al., 2015). These limitations stem from unavoidable causes including natural variation between students' rates of learning progress, and margins of error associated with the assessment scores themselves (Singer & Willett, 2003). Failure to account for these factors can result in spurious classifications and comparisons of progress for a non-trivial proportion of students. This paper argues that placing too much emphasis on individual progress metrics that are based on only two scores is likely to be counterproductive in practice. Instead, it is concluded that pairs of consecutive scores are best considered as part of a longer-term trajectory of scores along a clear progression of learning.

Defining gain and growth

Recommendation 4 of *The report of the review to achieve educational excellence in Australian schools* (Department of Education and Training, 2018) draws attention to the importance of gain and growth and invites clarification of the definitions of these terms:

> Introduce new reporting arrangements with a focus on both learning attainment and learning gain, to provide meaningful information to students and their parents and carers about individual achievement and learning growth (p. xiii).

Noting that terminology about learning progress can be varied (Hollingsworth et al., 2019), in this section we refer primarily to references that are concerned with quantification of progress using assessment scores. Assessment scores in isolation are sometimes called status measures (Castellano & Ho, 2013a). Terms like achievement and attainment are also used, as seen in the above recommendation. Moving beyond status to consider progress, it is generally accepted that progress measures require scores from the same student or students on multiple occasions. These serial data are referred to as longitudinal.

Contemporary research and practice on reporting progress using assessment scores reveals that many implementations are limited to quantifying progress using scores from only two successive occasions (O'Malley et al., 2011). Nese et al. (2013) and Ployhart and MacKenzie (2015) point out that this 'change score' between two occasions does not properly characterise growth, but instead would be more accurately characterised as gain. This seems like a useful distinction given the increased complexity of the statistical models that accommodate scores from more than two occasions and the more robust inferences about progress they can support (Curran et al., 2010).

The technical superiority of growth measures has at least two contributing factors. First, with the additional data points it is possible to average out or statistically account for measurement error and other statistical artefacts that plague simpler gain measures. Second, there is the capacity to construct and compare trajectories that contain nuanced information about growth by modelling change over time as a continuous process (Willett, 1994). Nonetheless, the naturally greater availability of gain information relative to growth trajectory information provides strong motivation to make use of the former whilst accommodating its limitations.

Preconditions for meaningful progress measurement

For gain or growth modelling that can meaningfully be related to learning in a given domain, the assessment should ideally have the following characteristics:

- all scale scores within a domain within the same assessment program should be on a common 'vertical scale' with interval properties
- each assessment already has, as part of its reporting framework, described proficiency levels that provide a criterion-referenced interpretation of progress.

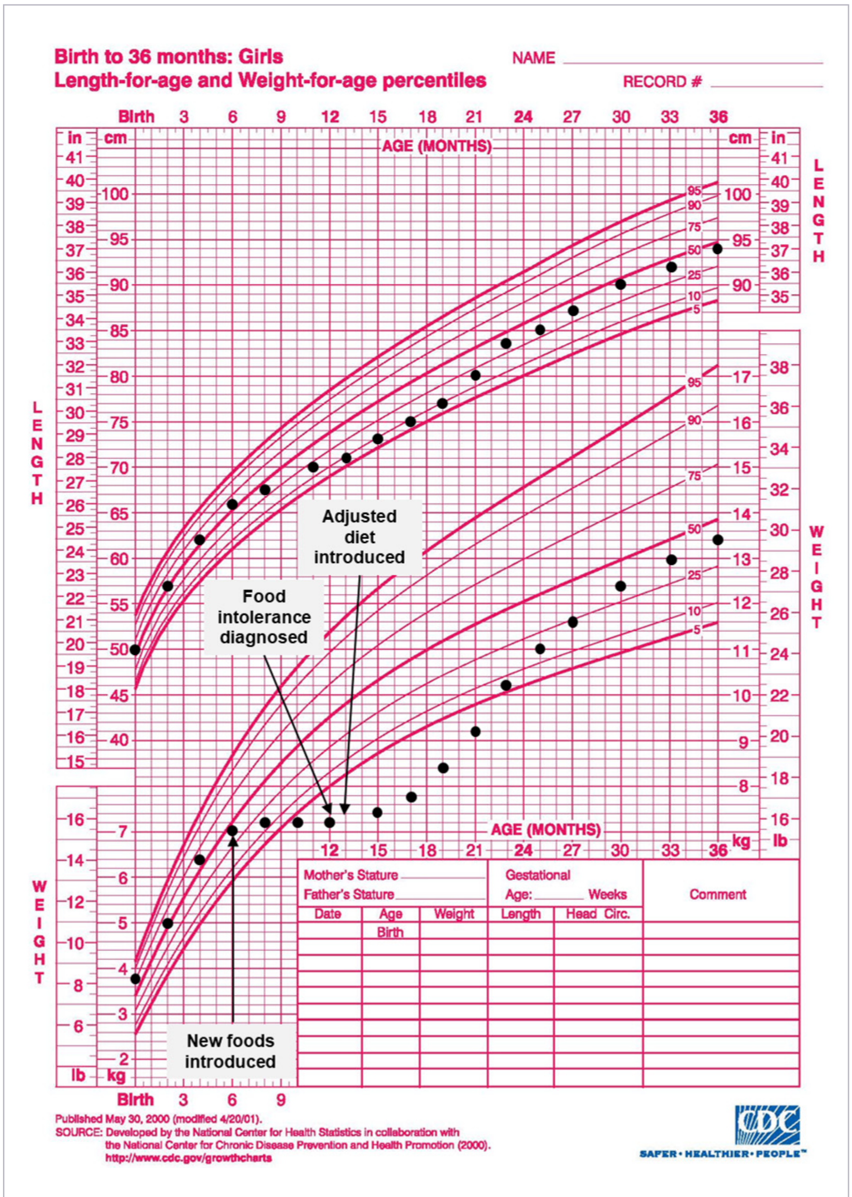
If these conditions cannot be met in practice, then there will be limitations on the range of methodological options for modelling and interpreting progress in a valid and meaningful way (Patz, 2007; Protopapas et al., 2016; Sireci et al., 2016).

What does learning growth typically look like on a scale?

This section summarises what learning growth often looks like when evidenced using assessment scale scores. This provides an important basis for contextualising changes in scale scores from one occasion to the next, particularly as they relate to making quantitative comparisons between the progress of individuals and groups. First though, it is instructive to consider growth against well-defined scales that measure attributes other than learning.

A well-known example comes from paediatric contexts, where measures such as the height, weight and head circumference of infants over time provide key developmental indices (see Figure 1). Much as in learning contexts, substantial deviations from typical trajectories can indicate that additional or different interventions are required. In these cases, the trajectory of growth following the intervention becomes of central interest. Parallels can be drawn with education, though successive measures from educational assessments are typically much more variable.

Figure 1 Example of a child's weight trajectory recorded on a paediatric growth chart



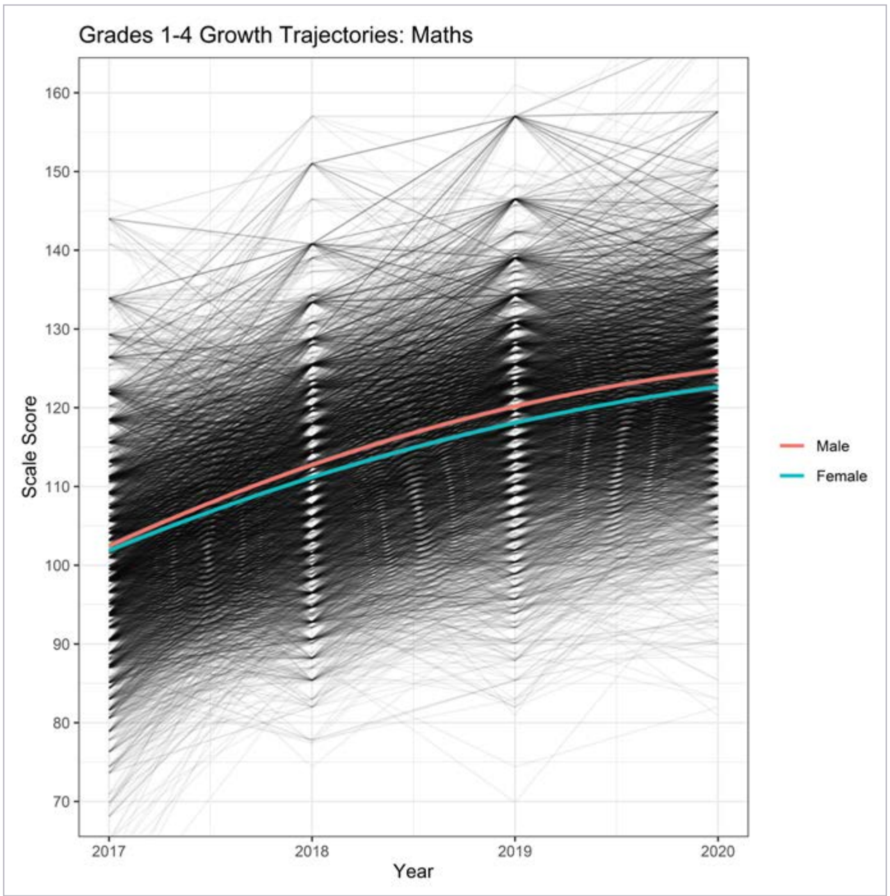
Growth chart template sourced from Kuczmarski et al. (2002, p. 36)

For many human attributes, evidence suggests that it is common for more substantial gains to be made initially. Growth rates often decelerate or stabilise with increasing amounts of the attribute. A brief scan of standardised assessment results and research literature from developmental psychology and school education contexts suggests that similar cohort-level patterns are commonplace (Australian Curriculum, Assessment and Reporting Authority, 2019; Li-Grining et al., 2010; Morgan et al., 2009; Williamson, 2018). However, this is not necessarily the case for all domains and age groups (Castellano & Ho, 2013a), and it is seldom true of every individual's growth trajectory.

Also of interest in educational research and evaluation is whether the growth trajectories of different groups of students differ. These groups may be categorised by contextual variables (e.g. school type), student characteristics (e.g. gender) or initial achievement levels (Singer and Willett, 2003). Whether the growth trajectories of different groups converge or diverge is also of key interest for detecting the so-called Mathew effect (Merton, 1968). In education this effect manifests as achievement gaps between groups that increase over time (Pfost et al., 2014).

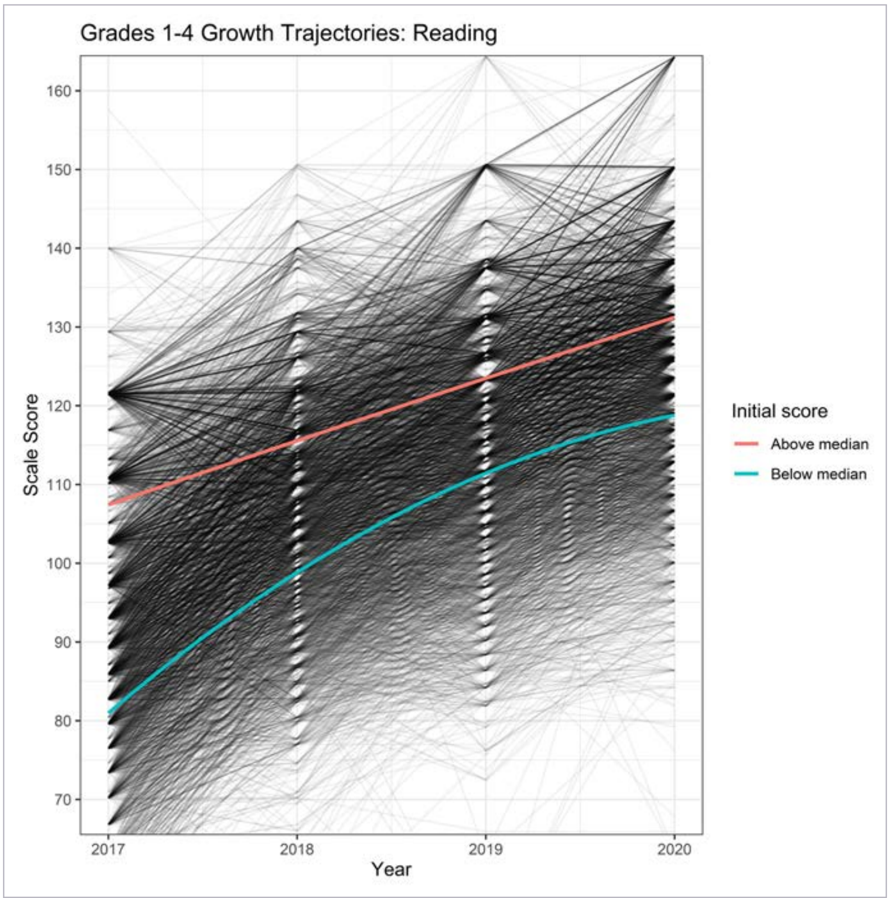
Figure 2 shows the average growth trajectories of female students and male students in a large sample of longitudinal Grades 1–4 PAT Mathematics data from Term 4 sessions. The grey lines in Figure 2 (sometimes referred to as a spaghetti plot) show how varied and volatile the observed initial scores and score gains can be for individual students. While it is difficult to visually discern, the volatility is greater among students with relatively extreme scores. After applying a three-level random intercept mixed-effects regression model using the lme4 (Bates et al., 2015) package in R (R Core Team, 2020), the following model parameterisation was well-supported: a quadratic (i.e. curvilinear) growth model fitted the data better than a linear growth model ($\chi 2(1, N = 20776) = 753.1$, $p = .00$); and, consistent with Figure 2, allowing the slope but not the intercept to differ across female and male students yielded the best model among several that were compared.

Figure 2 Individual and average growth trajectories for Years 1–4 PAT Maths for female and male students



The example in Figure 3 from PAT Reading shows initial convergence in the trajectories of students grouped by starting score (above or below the median) followed by more consistent growth rates, albeit at different levels. Looking at only the first year of progress, it is natural to conclude that the lower achieving group is making rapid progress in their learning and is on track to bridge the achievement gap, but there is a catch. The initially pronounced convergence observed here is in part a statistical artefact of having selected these groups on the basis of their initial scores. Grouping students in this way introduces an upward bias in the low scoring group, and vice versa, by inducing what is referred to as regression to the mean (Barnett et al., 2005). This phenomenon can have profound implications for interpreting scale score gains and is outlined in more detail in the following section.

Figure 3 Reading growth trajectories for students grouped by initial score above and below the sample median



Interpreting individual student gains

The discussion and patterns reviewed so far suggest the following tendencies:

- there is considerable variation in initial scores
- score gains can be volatile, particularly for students with extreme scores
- score gains sometimes taper off as students progress further up the scale
- score gains may differ between students grouped by certain characteristics.

The interpretation challenge is to take these observations into account when appraising an individual student's gain from one assessment to the next, and when making comparisons between the gains made by different students. The following two factors contribute to the observations just listed and have direct implications for interpreting gains:

- rates of learning (actual progress) vary across individuals and groups
- scores from all assessments contain measurement error.

These two factors introduce natural variation in the scale scores attained by students over consecutive occasions. This results in an imperfect level of correlation between initial scores and final scores, and imperfect correlations will always be accompanied by regression to the mean (Kahneman, 2011).

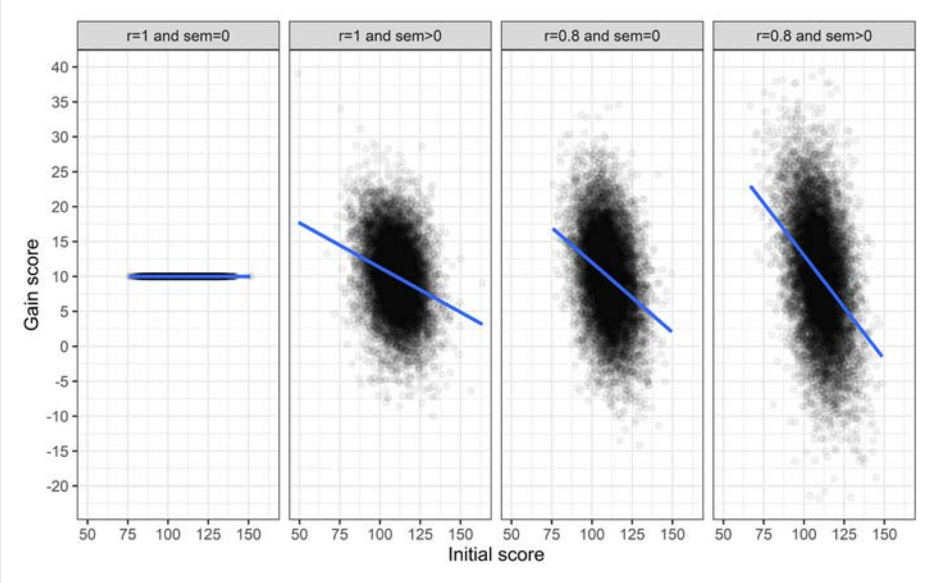
If on one measurement occasion random or idiosyncratic variation has a relatively large impact on a student's scale score, it becomes likely that on the second occasion it will contribute less. These statistical artefacts, particularly in a learning context characterised by genuine decelerating growth, produce the '…well-known negative correlation between prior score and gain …' (Betebenner & Linn, 2009, p. 6). In particular, students with prior scores higher than the population mean will systematically tend to show lower gains, and vice versa. Regression to the mean can make natural variation in repeated data look like real change.

Figure 4 illustrates this phenomenon by comparing the relationship between gain scores (i.e. final score minus initial score) and initial scores under different simulated conditions with the following parameters:

- population size of 10 000 students
- initial scale scores with a mean of 110 and a standard deviation of 10
- final scale scores with a mean of 120 and a standard deviation of 10
- latent or true correlation ('r') between initial and final scores of either 1 (i.e. all students gain exactly 10 scale scores) or 0.8 (close to that for PAT assessments taken one year apart after disattenuating for measurement error)
- measurement error ('sem': standard error of measurement) set at either zero (perfectly precise measurement) or between 3.5 and 6.5 following a quadratic error function giving extreme scores larger errors (errors are assumed to be uncorrelated).

The blue Loess fit lines in Figure 4 provide a moving average of the gain scores across the initial score range. These show that regression to the mean occurs as soon as there is measurement error in the assessment or as soon as there is an imperfect level of correlation between initial and final status. It is also clear that these two factors have a cumulative impact. Comparisons of score gains with the average would be biased between 0–3 scale score points across the middle 95 per cent of initial scores in the most realistic scenario (right-hand panel). The direction (positive or negative) is determined by whether the initial score was above or below the population mean of 110. Larger systematic bias is present as expected for students with more extreme initial scores. Comparing gains between students with initial scores either side of this range will be subject to biases exceeding half of the average population gain made in one year.

Figure 4 Four simulated gain scenarios illustrating regression-to-the-mean in absolute gain measures



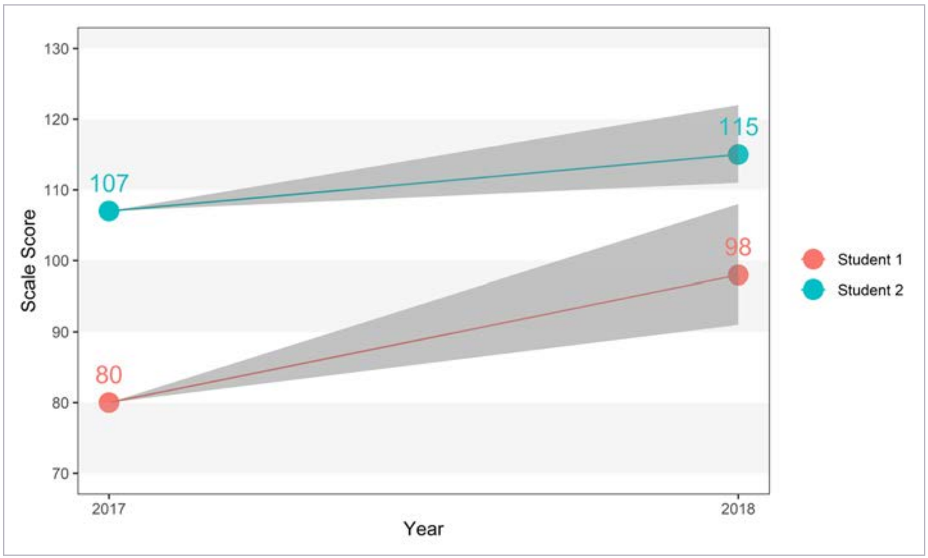
The simulation here has focused on revealing one unavoidable source of bias that can impact gain comparisons between students with different initial scores. Also worth noting is the unavoidable variation in gain scores resulting from the presence of realistic levels of measurement error. This can be seen clearly by examining the differences between the gain values in the first and second panels in Figure 4. The vertical spread of gain scores observed in the second panel but absent in the first panel is entirely attributable to measurement error. Depending on factors like the time between assessments, and the targeting of the assessments, this variation may be substantial enough to mask true gains or to mask biases due to regression to the mean. This is relevant for a non-trivial proportion of students, some of whom would attain a negative gain value due to chance alone. This matters when the focus is on quantifying, appraising and communicating individual student gains. Some researchers have argued that gain measures can be reliable when score distributions have certain characteristics (e.g. Rogosa et al., 1982; Williams and Zimmerman, 1996). However, we have observed that these characteristics are unlikely to apply to scale scores from high-quality, well-targeted assessments taken one year apart. The latter tend to more closely approximate distributional and correlational conditions known to be associated with low gain score reliability (e.g. Cronbach and Furby, 1970).

Statistical corrections can be made to account for regression to the mean in some situations (Rogosa et al., 1982), but this is not always practical or technically feasible. Therefore, in addition to expecting some volatility in absolute gain measures, anticipating asymmetries in gain scores across the initial score range is critical for ensuring that changes in scores are responded to proportionately. This in turn ensures that learners and educators are supported to direct their efforts in a targeted way. Being able to avoid incorrect conclusions about student progress, such as that a school appears to be doing a better job improving the learning of its lower achieving students than its higher achieving students according to gain scores alone, is one example of why these statistical considerations matter in practice.

A consequence of these biases is that absolute gains are often perceived as unfair for comparing the progress of individuals and groups who differ substantially in their prior achievement. To help contextualise whether an observed absolute gain is ‘typical’ or otherwise in the presence of these biases, it can be helpful to draw upon normative information. Several norm-referenced interpretations are possible, starting with simple comparisons to available cross-sectional scale score norms (like in Figure 1) and progressing to conditional metrics that take into account prior achievement and possibly contextual variables.

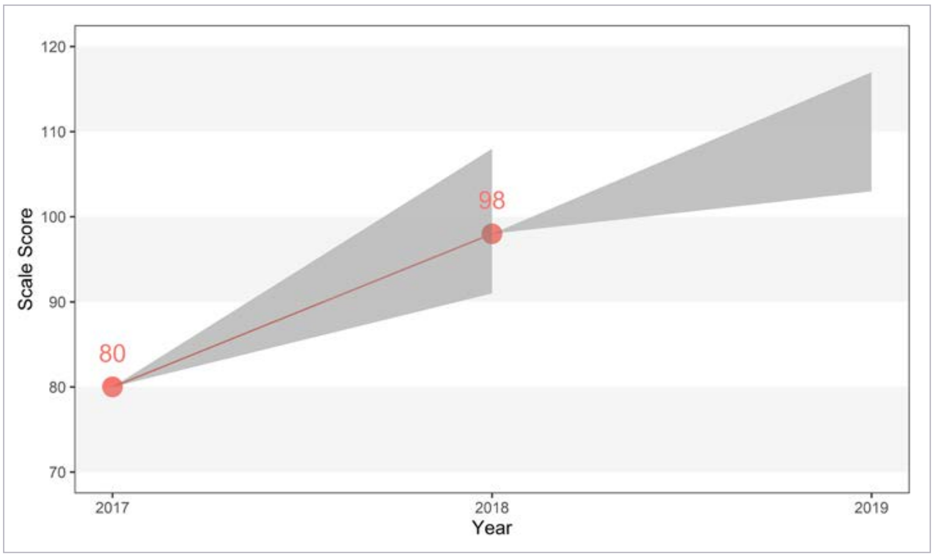
The use of models that compare progress between students with similar prior scores has emerged as a popular way to make these biases less visible through the construction of ostensibly fairer comparison groups for each individual student. A simplified version of this approach based on calculating absolute gain percentiles for students grouped by similar prior scores is shown in Figure 5. The middle 50% of these relative gain percentiles is shaded dark grey. It is worth noting that a variety of alternative calculation methods exist, including relative gain or conditional status measures (e.g. Castellano & Ho, 2013a, 2013b) and Student Growth Percentiles (SGPs) (Betebenner, 2011). Here we will use the term relative gain percentile since we adopt a simplified percentile-based calculation rather than a conditional regression-based calculation. It can be seen that the different levels of absolute gain for Student 1 (18 scale scores) and Student 2 (8 scale scores) both result in relative gain percentiles that are close to the middle of their respective relative gain distributions. The grouping by prior scores has ameliorated some of the biases that undermine comparisons between students who start at markedly different locations on the scale.

Figure 5 Comparison of gains for students with markedly different initial scores



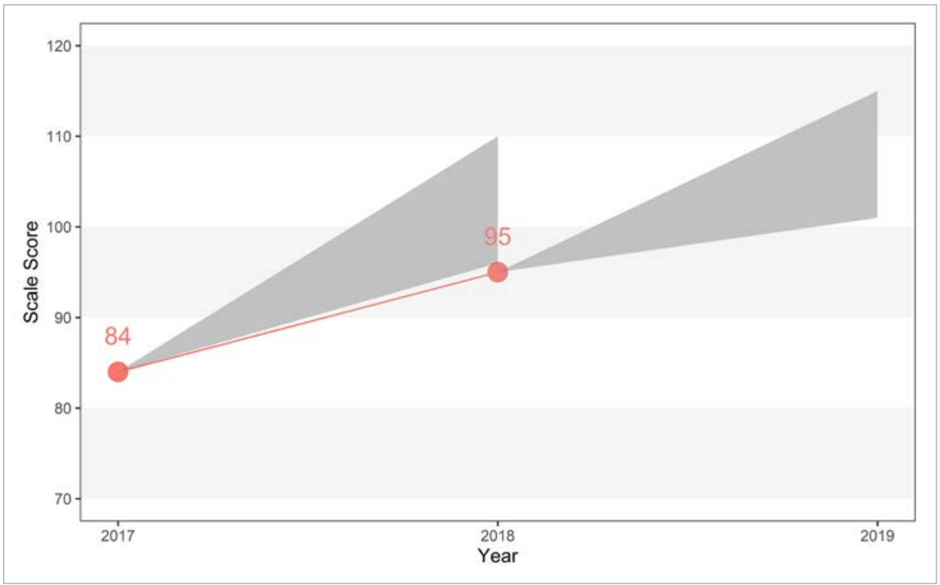
Using recent historical data from the same assessment, it is also possible to show projections of typical gain ranges that take prior score into account. Projections like these can be found in some reporting systems (Betebenner, 2011). The projection in Figure 6 shows the range of scale scores obtained historically by the middle 50 per cent of students who had also started with a scale score close to 98 one year prior. This kind of depiction may be useful for stimulating discussion and setting expectations about future learning goals and progress.

Figure 6 Relative gain percentile distribution for recent scores and as a projection



Unfortunately, the metrics that these relative or conditional models produce, if based on only one prior score, are volatile (McCaffrey et al., 2015; Sireci et al., 2016). For Student 1 and Student 2 in the earlier example, whose relative gains placed them close to the median, they could with non-trivial probability be classified as being in the lower or upper relative gain quartile after allowing for a realistic perturbation of scale scores by approximately one standard error of measurement (usually 3 to 4 scale score units). This is illustrated in Figure 7.

Figure 7 Illustration of classification volatility of relative gain percentiles



This simplified example is emblematic of the non-trivial levels of misclassification that can arise when using relative gain or conditional status metrics for individual students. From a measurement standpoint, there is little impact on the substantive interpretation of knowledge and skill for scale scores that have been perturbed within the bounds of measurement error. Correspondingly, the achievement bands would be stable or at most would change by one band near the level boundaries. In contrast, it is not unusual for standard errors associated with relative gain or conditional status percentiles to be as large as 15 percentile points (Sireci et al., 2016). In this situation, an estimated one-in-ten students with an observed relative gain percentile of 50 could be operating in the upper or lower conditional gain quartiles. This is consistent with modelling by Betebenner et al. (2016) who showed that approximately one-in-six students with an observed conditional percentile of 50 might in reality be below the 35th percentile progress benchmark used in that context. It follows that caution is required when interpreting individual student gain metrics like these and when using them to label the gains of individual students as typical or otherwise.

While conditional or relative gain approaches largely overcome comparability biases due to regression-to-the-mean and tapering growth trajectories, their apparent accentuation of measurement error is an unfortunate shortcoming. This brings into question the reliability of such metrics and the inferences made using them. These kinds of metrics are sometimes touted for diagnostic purposes, for example to identify students with relatively low gains who may need further support (Betebenner et al., 2016). However, even for this laudable purpose, some allowance for measurement error ought to be made or many false positives could arise.

The exposition so far on relative gain or conditional status metrics for individual students may seem disparaging. Nonetheless, these metrics can be helpful for understanding the range of gain scores that are historically 'typical' for students with similar prior performance in the given measurement context. The following conditions also go some way towards increasing the reliability of conclusions based on these metrics and might make reasonable preconditions for their adoption in practice:

- ensuring assessments are well-targeted for all individual students, for instance through adaptive assessment designs
- incorporating additional prior scores when constructing 'like groups' against which to compare gains
- triangulating other evidence about learning progress in the same domain.

These metrics are much less impacted by measurement error and therefore more reliable when aggregated across many students. However, even when aggregated, they are not completely free of bias and care should be taken in their analysis (Lockwood & Castellano, 2017).

So, what is there to gain?

Gain information is more readily available than robust growth trajectory information, but it is inherently volatile and subject to biases that complicate its use. These limitations beg the question of just how much weight to give to individual student gain metrics in practice, whether absolute or relative, for monitoring and responding to evidence about an individual student's progress.

Viewing the two consecutive scores as two of many along a longer-term progression of increasing knowledge and skill provides more solid footing. This is consistent with the growth mindset advocated by Masters (2016). This frame of reference could include described proficiency levels or learning progression levels or qualitative achievement standards. Given that each level or band occupies a scale score interval usually much larger than a standard error of measurement, these criterion-referenced or standards-referenced progressions provide much more stable markers of progress.

The availability of gain information from numerous assessments invites critical reflection. In the absence of more robust growth trajectory information, absolute gain measures and their normative derivatives might best be incorporated with caveats to augment substantive interpretations of individual student progress. Without this additional score information or this stable, longer-term frame of reference for learning progress, it seems there is little more to gain.

References

Australian Curriculum, Assessment and Reporting Authority. (2019). NAPLAN achievement in reading, writing, language conventions and numeracy: National report for 2019, ACARA.

Barnett, A. G., van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, *34*(1), 215–220. https://doi.org/10.1093/ije/dyh299

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Betebenner, D.W. (2009). Norm- and criterion-referenced student growth. *Educational Measurement: Issues and Practice*, *28*(4), 42–51.

Betebenner, D. W. (2011, 19 June). *New directions in student growth: The Colorado growth model*. Paper presented at the National Conference on Student Assessment, http://ccsso.confex.com/ccsso/2011/webprogram/Session2199.html

Betebenner, D. W. & Linn, R. L. (2009). *Growth in student achievement: Issues of measurement, longitudinal data analysis, and accountability*. Paper presented at the Exploratory Seminar: Measurement Challenges Within the Race to the Top Agenda.

Castellano, K. E., & Ho, A. D. (2013a). A Practitioner's guide to growth models. Council of Chief State School Officers.

Castellano, K. E., & Ho, A. D. (2013b). Contrasting OLS and Quantile Regression Approaches to Student 'Growth' Percentiles. *Journal of Educational and Behavioral Statistics*, *38*(2), 190–215. https://doi.org/10.3102/1076998611435413

Cronbach, L. J., & Furby, L. (1970). How we should measure change—or should we? *Psychological Bulletin*, *74*, 68–80.

Curran, P. J., Obeidat, K., & Losardo, D. (2010). Twelve Frequently asked questions about growth curve modeling. *Journal of Cognition and Development: Official Journal of the Cognitive Development Society*, *11*(2), 121–136. https://doi.org/10.1080/15248371003699969

Department of Education and Training. (2018). *Through growth to achievement: Report of the Review to Achieve Educational Excellence in Australian Schools*. Commonwealth of Australia. https://www.dese.gov.au/quality-schools-package/resources/through-growth-achievement-report-review-achieve-educational-excellence-australian-schools

Hollingsworth, H., Heard, J., & Weldon, P. R. (2019). *Communicating Student Learning Progress: A Review of Student Reporting in Australia*. Australian Council for Educational Research. https://research.acer.edu.au/ar_misc/34

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kuczmarski, R. J., Ogden, C. L., Guo, S. S., Grummer-Strawn, L. M., Flegal, K. M., Mei, Z., Wei, R., Curtin, L. R., Roche, A. F., & Johnson, C. L. (2002, May). *2000 CDC growth charts for the United States*. National Center for Health Statistics. https://www.cdc.gov/nchs/data/series/sr_11/sr11_246.pdf

Li-Grining, C. P., Votruba-Drzal, E., Maldonado-Carreño, C., & Haas, K. (2010). Children's early approaches to learning and academic trajectories through fifth grade. *Developmental Psychology*, *46*(5), 1062–1077. https://doi.org/10.1037/a0020066

Lockwood, J. R., & Castellano, K. E. (2017). Estimating true student growth percentile distributions using latent regression multidimensional IRT models. *Educational and Psychological Measurement*, *77*(6), 917–944. https://doi.org/10.1177/0013164416659686

Masters, G. N. (2016, 25 July). Monitoring student growth. *Teacher*. https://www.teachermagazine.com.au/columnists/geoff-masters/monitoring-student-growth

McCaffrey, D. F., Castellano, K. E., & Lockwood, J. R. (2015). The impact of measurement error on the accuracy of individual and aggregate SGP. *Educational Measurement: Issues and Practice*, *34*(1), 15–21.

Merton, R. K. (1968, 5 January). The Matthew Effect in science. *Science*, 56–63.

Morgan, P. L., Farkas, G., & Wu, Q. (2009). Five-year growth trajectories of kindergarten children with learning difficulties in mathematics. *Journal of Learning Disabilities*, *42*(4), 306–321. https://doi.org/10.1177/0022219408331037

Nese, J. F. T., Lai, C-F., & Anderson, D. (2013). *A primer on longitudinal data analysis in education. Technical report #1320*. Behavioural Research and Teaching. University of Oregon. https://files.eric.ed.gov/fulltext/ED545257.pdf

O'Malley, K., Murphy, S., McClarty, K., Murphy, D., & McBride, Y. (2011, September). *Overview of student growth models*. White Paper. Pearson.

Patz, R. (2007). *Vertical scaling in standards-based educational assessment and accountability systems*. Council of Chief State School Officers.

R Core Team (2020). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence.* Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195152968.001.0001

Sireci, S. G., Wells, C. S., & Keller, L. A. (2016). *Why we should abandon student growth percentiles.* (Research Brief No. 16-1). Center for Educational Assessment. http://www.umass.edu/remp/news_SGPsResearchBrief.html

Pfost, M., Hattie, J., Dörfler, T., & Artelt, C. (2014). Individual differences in reading development: A review of 25 years of empirical research on Matthew Effects in reading. *Review of Educational Research*, *84*(2):203–244. https://doi.org/10.3102/0034654313509492

Ployhart, R. E., & MacKenzie, W. I., Jr. (2015). Two waves of measurement do not a longitudinal study make. In C. E. Lance & R. J. Vandenberg (Eds.), *More statistical and methodological myths and urban legends* (pp. 85–99). Routledge/Taylor & Francis Group.

Rogosa, D., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin*, *92*(3), 726–748. https://doi.org/10.1037/0033-2909.92.3.726

Willett, J. B. (1994). Measuring change more effectively by modeling individual change over time. In T. Husen, & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (2nd ed.). Pergamon Press.

Williams, R. H., & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement*, *20*(1), 59–69. https://doi.org/10.1177/014662169602000106

Williamson, G. L. (2018). Exploring reading and mathematics growth through psychometric innovations applied to longitudinal data, *Cogent Education*, *5*(1), 1464424, https://doi.org/10.1080/2331186X.2018.1464424

Leading system transformation: A work in progress

Gregory B Whitby AM
Catholic Education Diocese of Parramatta

Maura Manning
Catholic Education Diocese of Parramatta

Dr Gavin Hays
Catholic Education Diocese of Parramatta

https://doi.org/10.37517/978-1-74286-638-3_11

Gregory Whitby is a Fellow of the Australian Council of Educational Leaders, a Fellow of the Australian College of Education, a Fellow of Western Sydney University and Adjunct Professor at the Universal Business School Sydney. Greg's mission to transform schooling for every child and young person is driven by a strong commitment to justice and equity. He continues to challenge accepted thinking about learning and teaching for the change education can bring for all of us. Greg has been awarded an Order of Australia (AM) and a Papal Knighthood in the Order of St. Gregory the Great.

Maura Manning is the Director Learning at Catholic Education Diocese of Parramatta. She is a nationally accredited teacher who has held leadership roles in the independent and Catholic education sectors. Her research interests include leading innovative learning design and creative teaching practice.

Dr Gavin Hays supports school teams as they navigate change in implementing contemporary learning pedagogies and inspecting their impact. This work focuses on implementing and sustaining pedagogical change in the areas of inquiry learning, project- and problem-based learning, flipped learning, retrieval practice and science, technology, engineering and mathematics (STEM). In 2019, he completed his doctoral study on developing a new measure for conceptual knowledge under Professor Dr Henk Schmidt at Erasmus University, Rotterdam, the Netherlands.

Abstract

Internationally, the COVID-19 pandemic has profoundly disrupted the education sector. While NSW has avoided the longer periods of remote learning that our colleagues in Victoria and other countries have experienced, we have nonetheless been provoked to reflect on the nature of schooling and the systemic support we provide to transform the learning of each student and enrich the professional lives of staff within our Catholic learning community. At Catholic Education Diocese of Parramatta (CEDP), a key pillar of our approach is to create conditions that enable everyone to be a leader. Following the initial lockdown period in 2020 when students learned remotely, we undertook an informal teacher voice piece with the purpose of engaging teachers and leaders from across our 80 schools in Greater Western Sydney to reflect on and capture key learnings. This project revealed teachers and leaders reported very high feelings of self-efficacy, motivation and confidence in their capacity to learn and lead in the volatile pandemic landscape. These findings raised the question: how do we enable this self-efficacy, motivation and confidence in an ongoing way? This paper documents the systematic reflection process undertaken by CEDP to understand the enabling conditions a system can provide to activate everyone to be a leader in the post-pandemic future and the key learnings emerging from this process.

Introduction

The call for educational transformation is echoing throughout the world as we approach our post-pandemic future. In Australia, reviews of curriculum offerings are underway at both state and national levels as well as an examination of the pathways through education and how the outcomes of these diverse pathways are documented in a meaningful way. While CEDP has felt the fierce urgency of this call well before the pandemic, the rapid change that occurred throughout 2020 has accelerated this work. Societal change in this past year has been undeniable and further changes to the economy and employment patterns will emerge. The nature of teachers' work changed dramatically when lessons moved online last year and this opened possibilities for collaboration that had not previously been explored. Prior to the pandemic there were already concerns about teacher workloads and this has been further explored in recent papers (Gallop, 2021). As the staffing concerns that have cast an ominous shadow in our schools in recent years are coming to an outright crisis, we need to consider new ways of working and reflect on the use of systemic resources to support the transformation of teachers' work and students' learning.

Traditionally, as a system of Catholic schools, CEDP operates within a theory of action that sees some finite resources and expertise held at the centre and shared across schools to build the capacity of school leaders and teachers to bring about high-quality learning experiences for students at the local school level. During the volatile periods of 2020, principals observed leadership emerging from all aspects of their staff and high degrees of collaboration as teachers responded to rapid change. During this highly dynamic period, the Executive Director observed powerful leadership emerging at the local community level as the principals responded to their unique communities rather than trying to implement a 'one-size-fits-all' centrally managed model. This unique pandemic situation challenged the traditional structures and processes that have defined the system.

This challenge presented an ideal opportunity to reflect on our existing theory of action and the espoused strategic pillars of CEDP to determine the agility of our structures and processes in responding to changes and challenges.

Four priorities guide the work of our Catholic system:

- mission is countercultural
- learning is owned by the learner
- equity is the norm
- everyone is a leader.

Informed by this mission and these overarching strategic priorities, there are three key priorities for the System Learning Team (SLT) who support learning and teaching in schools by working with teachers in learning spaces, in partnership with school leadership teams and sometimes directly with students:

- expand a culture of innovation
- personalise learning for leaders
- champion diversity and equity.

The SLT operates within the following theory of action that connects the four actions and their intended influence or impact:

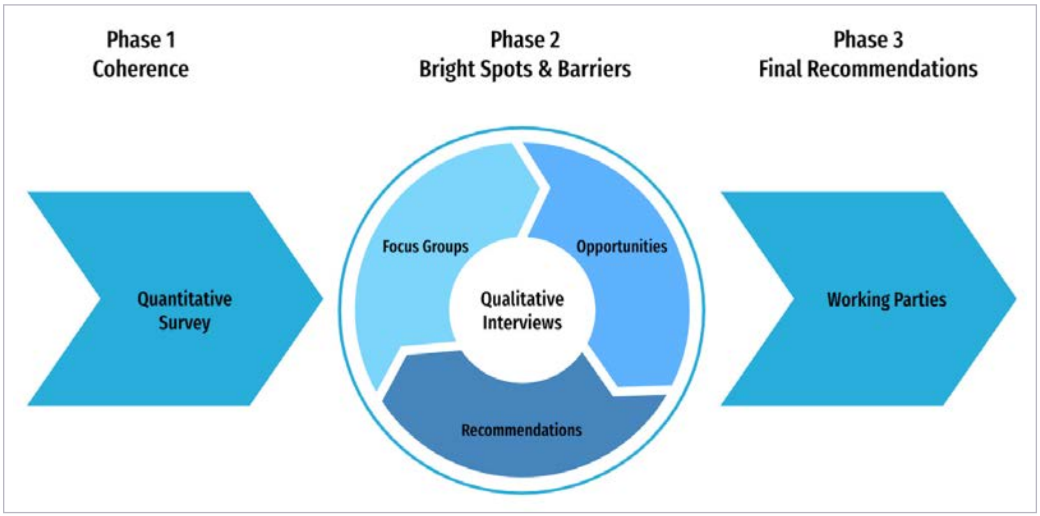
- teaching influences student learning
- teacher learning influences teaching
- school leadership influences teacher learning
- system leadership influences school leadership.

This paper documents a transparent and collaborative reflection on how the SLT supports schools to (in addressing the system intent) transform the learning of each student and enrich the professional lives of staff within our Catholic learning community.

Methodology

The approach to this work was iterative and the process has been collaborative and transparent. We designed a three-phase structure referred to as the 'learning health check' that began with a quantitative survey delivered to all teaching and non-teaching staff in schools (including leaders) as well as the SLT. This was Phase 1. The aim of the quantitative survey was to assess the coherence of our resources and services provided in meeting the needs of our schools. It is envisaged that this instrument could be used as an annual measure of coherence. Phase 2 involves a more qualitative lens in further exploring the key themes identified in Phase 1 by assembling focus groups that will address some of the areas of misalignment to identify opportunities and recommendations that may not have been previously considered. The future opportunities identified in the focus groups will be categorised according to implementation suitability and timeframe. Finally, Phase 3 will utilise working parties across school and system learning to co-design and test prototype solutions of new ways of working. Figure 1 shows the components of each phase.

Figure 1 Three phase structure for the 'learning health check'



Phase 1: Coherence

A quantitative survey was constructed to evaluate the following two areas:

- Impact – learning strategic priorities including:
  - impact on student learning (ISL)
  - culture of innovation (CI)
  - personalised learning (PL)
  - equity and diversity (ED)
- Process – learning theory of action
  - student learning (SL)
  - teaching (TH)
  - teacher learning (TL)
  - school leadership (SL)
  - system leadership (SYL)

The survey questions were designed to collect insights regarding the areas of impact and process across different roles and contexts within school-based personnel and SLT. Table 1 shows the questions that were used in the design of the qualitative survey.

The quantitative survey was administered in December 2020 and was distributed to both school-based personnel and SLT.

Table 1 Questions used to address the design of the qualitative survey

	Impact on student learning (ISL)	Culture of innovation (CI)	Personalised learning (PL)	Equity and diversity (ED)
Question stem: The learning directorate:				
Student learning (SL)	1.1 Supports me to improve the learning outcomes of each student.	2.1 Is open to new ideas	3.1 Provides choice in how I solve student learning problems.	4.1 Provides access to the resources and services that I need to improve the learning outcomes of each student.
		2.2. Values the contributions and insights of teachers to improve student learning outcomes of each student.		
Teaching (TH)	1.2 Provides support in developing teaching practices that improve the learning outcomes of each student.	2.3 Provides in situ support that enables me to take ownership over new teaching practices that improve the learning outcomes of each student.	3.2 Provides a range of in situ learning opportunities to develop teaching practice that are personalised to my needs.	4.2 Listens and understands my diverse needs.
		2.4 provides resources that enable me to take ownership over new teaching practices that improve the learning outcomes of each student.		4.3 is collaborative in sourcing appropriate support and resources to improve the learning outcomes of each student.
Teacher Learning (TL)	1.3 provides professional learning opportunities that assist me to transform the learning of each student.	2.5 provides me with new knowledge and skills that enrich my professional life.	3.3 helps identify learning and teaching needs within my school.	4.4 is responsive to school needs in the resource and service allocation to build the capacity of all teachers.
			3.4 provides support to address these needs with fair, constructive feedback on my work.	

	Impact on student learning (ISL)	Culture of innovation (CI)	Personalised learning (PL)	Equity and diversity (ED)
School Leadership (SL)	1.4 encourages school leaders to work collaboratively to lead the learning agenda within their school context to transform the learning of each student.	2.6 introduces innovative ideas into work practices that enrich the professional lives of staff.	3.5 provides personalised opportunities for leaders to transform the learning of each student.	4.5 provides autonomy in the management of the resources and services that schools access to transform the learning of each student.
System Leadership (SYL) The learning directorate:	1.5 works in partnership with schools to provide mutual support that identifies, understands and scales improvement efforts that transform the learning of each student.	2.7 champions innovation by seeking out new strategies or instruments that support schools to transform the learning of each student.	3.6 facilitates diverse professional learning opportunities that enrich the professional lives of staff.	4.6 promotes opportunities for joint decision-making in the allocation of resources and services with each school context.
	1.6 works in partnership with schools to provide an appropriate level of challenge that identifies improvement efforts that transform the learning of each student.	2.8 champions innovation by seeking out learning frameworks that support schools to transform the learning of each student.	3.7 supports diverse professional learning opportunities that enrich the professional lives of staff	4.7 promotes opportunities for divergent opinions in the allocation of resources and services with each school context.

Phase 1 response rates

Figures 2 shows the survey response rate of the total 971 school-based personnel responses. And Figure 3 shows the survey response rate of the total 70 system-based personnel responses. Descriptive statistics were calculated to identify any significant differences between the identified subgroups. Some of which included school type (primary and secondary), school population, leadership role (senior and other leader) and system team roles. Within the survey construction, qualitative responses were also collected and coded using Nvivo to identify possible themes for further investigation in Phase 2. In March 2021, a summary of findings was presented to system leaders and they were given the opportunity to engage their school-based personnel in an opportunity to ask clarifying and probing questions related to the findings. This process again was an opportunity to continually build relational trust within the system and an opportunity to co-construct next steps.

Figure 2 Survey response rate for school-based personnel responses

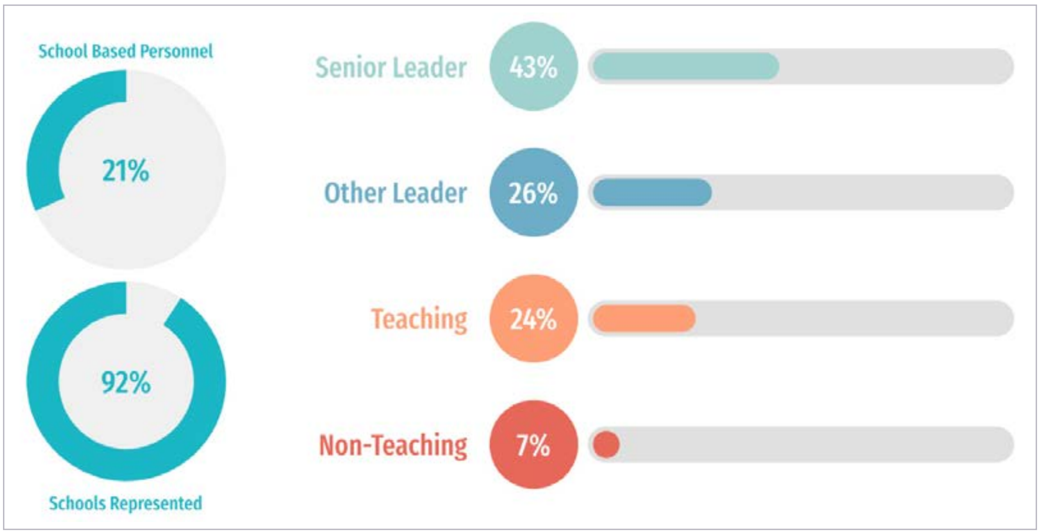
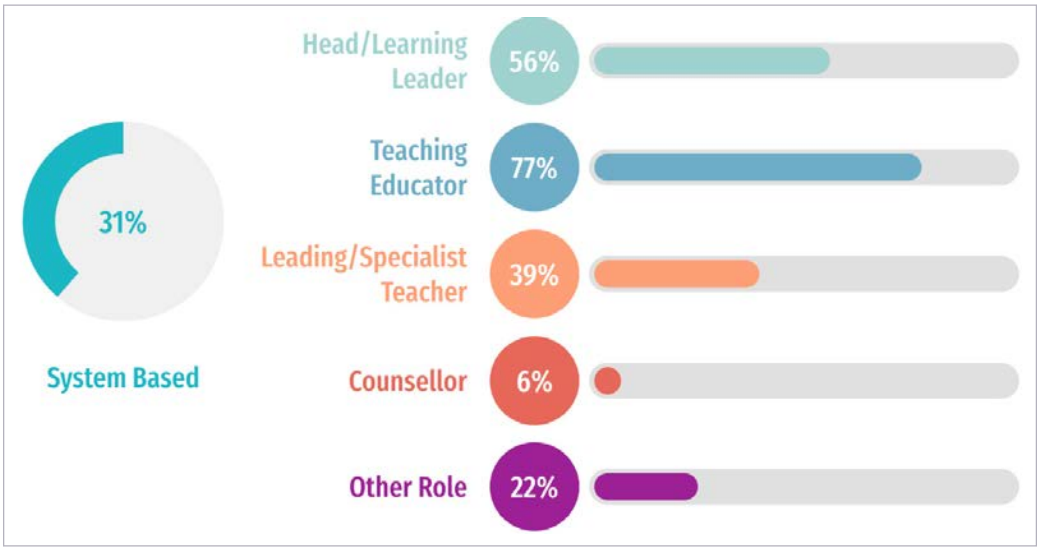


Figure 3 Survey response rate for SLT responses



Priority areas

From the survey in Phase 1, we identified three priority areas that merited further investigation or attention due to the lower ratings or the discrepancies between school-based personnel and SLT responses.

- Priority 1 Alignment and trust between SLT and schools to enable leadership in all.
- Priority 2 Importance of structures and processes to support diversity and equity.
- Priority 3 Ensuring structures and processes foster innovation and transformation.

This paper focuses on Priority 1.

Priority 1   Alignment and trust between SLT and schools to enable leadership in all

The quantitative data identified a pattern around the misalignment of perceptions between SLT and school-based personnel. SLT staff reported higher perceptions of their effect than school-based personnel. The qualitative comments flagged some concerns around the collaborations between the system office and schools. The responses suggest there may be weak relational trust between the system office and schools in some instances.

Some qualitative responses suggested some respondents did not share a belief in the Catholic systemic principle of making better use of finite resources by sharing expertise across schools. However, the feedback also suggested that when learning or other services are personalised or tailored to the context, respondents were most positive. It was apparent that when respondents felt their context was deeply understood by system office-based staff they felt their interactions were more positive.

This customised service is linked to respondents' desire to have genuine agency in all that they do. Respondents suggested that 'one-size-fits-all' approaches and policies or programs that eroded schools' capacity to make decisions based on their contexts left them feeling frustrated. Some respondents reported feeling constrained by the perception that system office personnel sometimes imposed a single way of doing things on schools without listening or understanding the context. There was a perception that the system office staff had 'an agenda' for their work that was not always in keeping with what the school believed was their priority.

We sought to explore how to build greater coherence and overall collective efficacy (Donohoo et al., 2018) as a learning system to activate all members of the community to transform the learning of each student and enrich the professional lives of staff through a series of focus groups with teachers, leaders and office-based staff.

Phase 2: Bright spots and barriers

The key insights from the quantitative data and the coding of the qualitative responses were grouped to focus on three priority areas. These were further explored through a series of targeted focus groups that invited participants to imagine possibilities for the future and ultimately engage colleagues in a co-design and testing of prototype solutions. The facilitation of the focus group meetings utilised a 'step in, step out and step back' protocol to enable participants the opportunity to identify bright spots in the way we currently work that make a difference in their role or context. But also, to explore the barriers or areas of possible misalignment. This protocol required participants to analyse the impact locally and broadly across the system to schools to ideate possible prototype solutions that could be deepened in Phase 3. Phase 3 is currently being undertaken at the time of writing this paper..

Conclusions

While this remains a work in progress at the time of submission of this paper, some key themes have emerged through the focus group ideation process that will inform the future processes and structures of our system.

1. The constant transparent process of building trust with stakeholders through co-construction and collaborative unpacking is essential. Often at a system level,we expedite processes to get to outcomes.
2. System structures established to support improvement are not the same structures that support innovation and transformation.
3. A linear theory of action suggests knowledge, expertise and power is held by 'authority figures', which may contribute to erosion of empowerment, creativity and learning of teachers.
4. System structures must enable ideas and expertise to emerge from any part of the organisation to enable anyone to be a leader.
5. An implementation stance erodes teacher agency and learning capacity.
6. The complexity of needs in classrooms demand new ways of working and new solutions.
7. Teacher learning must be central to a teacher's day, not an add-on.
8. Horizontal connections between schools enable teachers to build system knowledge and contribute new understandings to the system discourse.

Following this idea generation phase, we will move to identifying the best possible solutions. Different stakeholder groups will be invited to engage with these ideas and participate in the discernment process. Through this process, we will seek to identify potential 'prototypes' that can be trialled and evaluated to inform our processes and structures in our post-pandemic future.

References

Donohoo, J., Hattie, J. & Eells, R. (2018). The power of collective Efficacy. *Educational Leadership*, *75*(6), 40–44.

Gallop, G., Kavanagh, T. & Lee, P. (2021). *Valuing the teaching profession: An independent inquiry*. NSW Teachers Federation. https://www.nswtf.org.au/pages/reports

Making excellent progress in early reading: How can the identification of essential skills and concepts help?

Dara Ramalingam, Prue Anderson, Sandra Knowles, Danielle Anzai and Greta Rollo
Australian Council for Educational Research

https://doi.org/10.37517/978-1-74286-638-3_12

Dr Dara Ramalingam is a Senior Research Fellow in the Assessment and Psychometric Division at ACER, where she has worked since 2004. Dara has worked extensively on assessments of 21st-century skills included in large-scale assessments run by the OECD. In the 2009 Programme for International Student Assessment (PISA), she played a leading role in test development for both paper-based reading, and digital reading, the latter being the first time digital reading was assessed in a large-scale way. Dara also contributed to the development of assessment material for the OECD Programme for International Assessment of Adult Competencies computer-based assessment of literacy. In PISA 2012, Dara oversaw assessment development for the computer-based problem solving material by coordinating the efforts of the four test development centres located in Australia, Germany, Norway and the United States. She also contributed to development of the entirely new PISA problem-solving framework for the 2012 cycle. More recently, Dara has been instrumental in the development of ACER's learning progressions, and is contributing to a project aiming to define and measure critical thinking, creativity, and collaboration.

Abstract

The ability to read and understand text is fundamental to full participation in modern adult life (Olson, 1977; Elwert, 2001). It is essential to educational progress across domains, but increased literacy levels are also linked to positive outcomes in terms of employment and health. Given its critical role both in the facilitation of learning in all domains, and in many aspects of life beyond school, it is imperative that we give students the best possible chance to develop their reading skills. This paper uses early reading as a case study for examining how the identification and explication of essential skills and concepts might assist all students to make excellent progress.

Why should we identify essential skills and concepts?

One of the major findings of a recent review of the NSW curriculum (Masters, 2020) was that many syllabus documents are overcrowded and that, because of time constraints, important ideas can often only be taught a superficial way. If the criticism that many current syllabuses are overcrowded to the point where important ideas, although present, can be obscured, then one solution would be to clearly identify and articulate ideas that should be prioritised in the classroom. That is, making explicit the essential concepts and skills of a learning area means that peripheral content need not be a focus.

In addition to providing a way of addressing the issue of overcrowding, the identification of essential skills and concepts forces one to be clear in defining them: what are they, and what is their justification? How do they relate to the learning area as a whole? Can we articulate the key ways in which growth occurs in these essential concepts? The remainder of this paper is dedicated to offering answers to these questions in relation to early reading.1

What is reading?

Reading requires a broad variety of perceptual, linguistic, and cognitive skills to extract meaning from visually presented material, most commonly, written text. In light of this, proficient reading takes years to develop and involves both understanding a language (comprehension) and understanding the symbolic representation of that language as written text (learning to read aloud).

A sequence of studies from the United States, notably Snow et al's., *Preventing reading difficulties in young children* (1998) and the National Research Council's *Starting out right: A guide to promoting children's reading success* (1999) culminated in an influential report published by the US National Institute of Child Health and Human Development, *Report of the National Reading Panel: Teaching children to read* (National Reading Panel, 2000). This comprehensive and seminal study, still relevant more than 20 years after its initial publication, found that a constellation of skills and knowledge is required to read, including 'the alphabetic principle, reading sight words, reading words by mapping speech sounds to parts of words, achieving fluency, and comprehension' (Snow et al., 1998, p. 6). Similarly, the *Teaching reading: Report and recommendations*, published in 2005, found that all students learn best when teachers adopt an integrated approach to reading that explicitly teaches phonemic awareness, phonics, fluency, vocabulary knowledge and comprehension (Rowe & National Inquiry into the Teaching of Literacy [Australia], 2005).

The development of proficiency in reading comprehension is broadly similar across languages. In all languages, proficiency in reading comprehension initially develops in relation to texts that are read aloud to learners. Later, these skills are applied (and further developed) in relation to texts that are read independently. So too in all languages, reading comprehension begins with the initial realisation that text contains meaning, passes through the capacity to understand short written texts presenting familiar ideas, and moves on to the capacity to understand and critically reflect on a broad range of sometimes long and complex written texts with layers of subtle meaning that present unfamiliar ideas and draw upon a wide vocabulary.

In the early years in particular, it is acknowledged that the development of reading comprehension is underpinned by other skills. For example, prior to being able to read and write independently, the only way that students can demonstrate their comprehension is through their oral language skills. As Castles et al. said, 'the foundation of reading comprehension is provided by oral language' (2018, p. 38). It follows that support for oral language development that is focused on quality written texts provides the foundation for the development of reading comprehension. Similarly, the development of vocabulary and general knowledge are also integral to the development of early reading.

> There is now widespread agreement that learning vocabulary, developing oral language skills, and acquiring knowledge are tasks to be tackled in early childhood and primary settings, and that they are just as important as are the tasks of learning letters, sounds, decoding, and fluency (Snow, 2017, p. 8).

While skills such as oral language development are clearly important in supporting the development of early reading comprehension, the focus here is on essential skills and concepts unique to reading.

1 English in the Australian National Curriculum is broad in its scope, covering language, literature and literacy. These areas are all critical, but this paper focuses on reading, specifically on identifying and describing the key skills that students need in order to start reading independently with comprehensive understanding – the comprehension of written text.

The essential skills and concepts of early reading

It is important to note that the essential skills and concepts outlined here are not new ideas: in fact they find strong expression both in existing curricula, as well as in the research literature (see for example, Snow, 2017; Castles et al., 2018; Freebody & Luke, 1990). The essential concepts are an attempt to identify the most fundamental ideas in early reading. While these essential concepts have strong support, they are sometimes difficult to identify in existing materials, particularly in the case of curricula and syllabus documents, which tend to outline a great deal of detailed content, without identifying which ideas are critical. In addition, much of the existing work does not adequately address the description of growth in the concepts identified. In an attempt to address this limitation, as well as the rationale for the essential concepts, a description of the beginning levels for one concept, text form and purpose, follows the introduction of the essential concepts.

Table 1 shows the five essential skills/concepts of early reading that have been identified through a review of existing bodies of work. Although their explanation is beyond the scope of this paper, the key aspects of each are included. An explanation of each skill/concept follows the diagram.

Table 1 The essential skills and concepts of early reading

1 **Representation and fluency**	2 **Text form and purpose**	3 **Critical perspectives**	4 **Interpreting meaning**		5 **Searching**
Segmenting, blending and manipulating sounds Phoneme-letter mapping Conventions of print Spelling Punctuation Fluency, text complexity and vocabulary	Text types and purposes Text structure and organisation Metalanguage for text forms and parts of speech	Ethical perspectives Evaluating text quality Logical analytic skills Reputational evaluations Recognition of implied bias	**Listening to and discussing texts** Text comprehension (see independent reading) Using the vocabulary, syntax and grammar of written texts	**Independent reading** General knowledge Making connections Identifying key ideas Comparing and contrasting Summarising and generalising Making inferences Identifying supporting evidence Recognising the intended effect of words Recognising lost meaning	Search criteria Search strategies Competing information

1. Representation and fluency

Students need to know how spoken words are represented in writing and how to quickly and accurately automatically recognise them (Snow, 2017; Castles et al., 2018). Integral to this is the understanding that writing 'says something', that it is the representation of the words they say and hear and therefore makes meaning. They need to know the print conventions used for directionality, punctuation, referencing and text formats. Most basic decoding skills of segmenting, blending and phoneme-letter mapping must be mastered in order to support fluency development and writing with error-free spelling (Castles et al., 2018). Fluency, which requires fast, accurate word recognition and appropriate word grouping, is essential to support independent reading comprehension. Fluency has a 'transformational impact' on reading: 'it is the point at which component skills are so automated and highly integrated that maximum cognitive energy is available to focus on meaning' (Konza, 2014, p. 161). Learners must become increasingly fluent readers who are able to automatically recognise a wider range of words as texts become increasingly complex.

2. Text form and purpose

Text form and purpose, is an important concept throughout literacy learning. Learners need to recognise the different forms and purposes of texts as this supports interpreting, searching or critiquing the text as well as the creation of different types of texts. Initially there is much to learn, as text forms and purposes are complex and varied. Once reasonably skilled readers and writers are familiar with a wide range of text types they will be able to navigate these texts, or emulate key features, developing increasing sophistication in their understanding as texts become more complex. Learners also need to develop increasingly nuanced metalanguage to describe and critique text forms including text structure, organisation and the way language is used as well as applying their understanding of text forms and purposes to their own text creations. Freebody and Luke's model (1990) encompasses a similar concept in its reference to 'text user', which is about understanding text types in the context of their purposes, or 'what the text is for, here and now' (p. 10).

3. Critical perspectives

Students need to become critical readers able to evaluate texts they read and justify their evaluations. For young readers, initial evaluations of texts they read are likely to be based on personal enjoyment or relate to elements of the text that present familiar experiences, or familiar values. As students mature and become more skilled and experienced readers, they learn how to adopt a range of critical perspectives, which eventually develop into a highly complex and demanding set of thinking and analytical skills that can be applied across a range of perspectives. These will extend into postgraduate tertiary education and highly specialist reading and writing tasks. Freebody and Luke's model encompasses a similar concept in its reference to 'text analyst', which they describe as an extension of critical thinking, 'an awareness of the fact that all texts are crafted objects, written by persons of particular dispositions or orientations to the information, regardless of how factual or neutral the products may attempt to be' (Freebody & Luke, 1990, p.13).

4. Interpreting meaning

Through listening to and discussing texts

When they are first learning to read, students can demonstrate more advanced skills in interpreting the meaning of texts when texts are read to them. This is partly because the short, simple texts used to master decoding are not designed with complexities in comprehension in mind. Therefore interpreting meaning through listening to and discussing texts underpins the development of reading. In early literacy, comprehension skills are developed through listening to and discussing texts and these skills gradually transfer to texts students can read themselves as they achieve sufficient fluency.

When reading independently

Readers need to construct much of the meaning as they read texts. Writers make assumptions about the prior knowledge of their readers. They do not explain everything. Aspects of the meaning may be assumed or implied. Even when meaning is explicit, more complex texts often require comprehension skills that go beyond simply knowing the vocabulary and understanding the grammar and syntax to understand the meaning. Readers need a broad range of comprehension strategies including skills that support forming a broad understanding as well as skills that support close interrogation. Reading comprehension is the coordination of a number of integrated processes (Castles et al., 2018; Conley & Wise, 2011; Kendeou et al., 2016).

5. Searching

Searching is specific to reading. One purpose of reading within a text is to locate a specific piece of information. Konza refers to searching as part of the purposeful approach of good reading practices, as good readers can, for example 'skim over a newspaper article lightly' or 'scan a page quickly for a telephone number' (Konza, 2014, p. 163). Searches may be conducted within texts, or across texts, to locate a specific piece of information. Search criteria may need to be developed or refined to begin the search. Some searches may need to meet multiple criteria. Knowledge of how indexes, glossaries, and search engines work may be required. Additional research skills are generally necessary to select reputable links in an internet search. Scanning within a text can be supported by effective use of headings, subheadings and other structural elements. Similarly, competing information that does not match the search criteria must be rejected. A text, selected from many, also needs to be quickly reviewed to identify whether it meets the intended purpose with similar, but less appropriate or irrelevant texts rejected.

The importance of descriptions of growth

Students in most Australian classrooms differ widely in their levels of proficiency. A body of evidence converges on the conclusion that the most advanced 10 per cent of students in each year of school are about five to six years ahead of the least advanced 10 per cent of students (see for example Siemon et al., (2019)). This is not a new problem, and teachers are well aware of the need to differentiate in their instructional activities. However, under the existing grade-based approach to teaching, syllabus documents suggest that all students in a given grade be taught the same syllabus. Given the broad range of different levels of proficiency present in a classroom, if the syllabus were followed exactly, some students would be taught material that is beyond their current level, while others would be taught material they have already mastered. The key to ensuring that students make progress is the identification of where students are in their learning according to the essential skills and concepts, and to target teaching accordingly. In order to make this identification, however, it is necessary first to clarify and understand what it means to develop proficiency in a learning area. So, as well as identifying the essential skills and concepts within a domain, (as has been attempted for early reading in the previous section), it is necessary to also clarify what growth in these essential skills and concepts might look like. Table 2 gives descriptions of several levels of the essential concept of text form and purpose. These descriptions are evidence-based, developed from valid and reliable assessment data that have identified a 'typical' trajectory of reading comprehension development. Therefore they can provide teachers with confidence in the data they are using to target areas of learning, and to identify how students' progress over time.

Table 2 Levels of text form and purpose

Level	Text form and purpose
Level 1	Use contextual clues such as location and images, to identify familiar book/materials (no word reading skills at this level).
Level 2	Recognise the purpose of a few highly familiar texts (e.g. labels, signs, well-known books read aloud for enjoyment).
Level 3	Listen to short, simple texts and orally identify obvious features (e.g. how likely the events in a story are) when clues are clear. Explain the purpose of very familiar visual conventions in illustrated texts (e.g. thought bubbles). Recognise the purpose of a few familiar texts read aloud when this is very obvious (e.g. to tell a story; sing a song; find a character in different images). Note this depends on exposure to these texts.
Level 4	Recognise the purpose and prominent elements of the form of a range of familiar texts read aloud (e.g. stories, information texts, recipes, lists, phone texts). Note this depends on exposure to these texts.

Conclusion

A common criticism of modern curricula is that they are overcrowded and focus on breadth of content rather than ensuring deep understanding of essential ideas. The identification of essential skills and concepts offers a way to add focus to existing curricula structures by encouraging thoughtful reflection of how different pieces of content might be taught in the context of one or more overarching fundamental concepts. To accompany the descriptions of growth in relation to the essential concepts, instructional materials are under development. These materials will be targeted at conceptual levels instead of school-year levels to demonstrate to teachers how explicit articulation of the early reading essential skills and concepts can help determine where their students are in their learning and target their teaching accordingly, an essential element in ensuring all students make excellent progress.

References

Castles, A., Rastle, K., & Nation, K. (2018). Ending the reading wars: Reading acquisition from novice to expert. *Psychological science in the public interest*, *19*(1), 5–51. https://doi.org/10.1177/1529100618772271

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

Elwert, G. (2001). Societal literacy: Writing culture and development. In D. Olson and N. Torrance (Eds.), *The making of literate societies* (pp.54–67). Blackwell.

Freebody, P. & Luke, A. (1990). Literacies programs: Debates and demands in cultural context. *Prospect: An Australian Journal of TESOL*, *5*(3), 7–16.

Konza, D. (2014). Teaching reading: Why the 'Fab Five' should be the 'Big Six', *Australian Journal of Teacher Education*, *39*(12). http://doi.org/10.14221/ajte.2014v39n12.10

Masters, G. N. (2020). *Nurturing wonder and igniting passion, designs for a new school curriculum: NSW Curriculum Review.* New South Wales Education Standards Authority.

National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read*. US Government Printing Office. https://www.nichd.nih.gov/sites/default/files/publications/pubs/nrp/Documents/report.pdf

National Research Council. (1999). *Starting out right: A guide to promoting children's reading success*. The National Academies Press. https://doi.org/10.17226/6014

Olson, D.R. (1977). The language of instruction: The literate bias of schooling. In R. Anderson, R. Spiro and W. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 65–89). Erlbaum.

Rowe, K., & National Inquiry into the Teaching of Literacy (Australia). (2005). *Teaching reading: Report and recommendations.* Department of Education, Science and Training. https://research.acer.edu.au/tll_misc/5

Siemon, D., Barkatsas, A., & Seah, R. (Eds.). (2019). *Researching and using progressions (trajectories) in mathematics education*. Brill Sense.

Snow, C. E. (2017). Early literacy development and instruction: An overview. In N. Kucirkova, V. Grøver & C. McBride (Eds.), *The Routledge international handbook of early literacy education: A contemporary guide to literacy teaching and interventions in a global context.* (pp. 2–19). Taylor & Francis.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. National Academy Press.

Supporting science teaching practice with learning progressions

Professor Erin Furtak
University of Colorado, Boulder

https://doi.org/10.37517/978-1-74286-638-3_14

Erin Marie Furtak, PhD, is Professor of STEM Education and Associate Dean of Faculty at the University of Colorado, Boulder. A former high school science teacher, Erin transitioned into a career studying how science teachers learn and improve their daily classroom practices through formative assessment. In a series of multiple studies, Erin has been partnering with teachers, schools and districts to learn how teachers can design, enact and take instructional action on the basis of classroom assessments that they design. Her recent publications have examined the ways in which the design and enactment of classroom assessments can promote more equitable participation in science learning.

Abstract

Learning progressions are often used as foundations for curriculum and assessment. At the same time, as representations of the development of student ideas and practices, they can also serve as maps to support teachers during instruction. This paper describes a program of research in which my colleagues and I have investigated how learning progressions can support high school science teachers in cycles of co-designing formative assessments.

Introduction

For decades, science education has emphasised students engaging in the thinking processes of scientists as they learn the big ideas in science. These reforms have been reflected in frameworks for science learning internationally, such as the Next Generation Science Standards in the US (National Research Council [NRC], 2012) and in the framework for the International Program for Student Assessment (Organisation for Economic Co-operation and Development [OECD], 2017). These ambitious visions for science learning raise expectations for the ways in which teachers design and facilitate learning environments that reposition student thinking at the centre of classroom activity and place high demands on teachers to enact. To realise this new vision of science education, teachers will need to participate in long-term, ongoing professional learning that provides them with tools and frameworks to navigate student thinking in three-dimensional learning experiences (National Academies of Sciences, Engineering, and Medicine, 2015).

In this paper, I describe an ongoing program of research in which I have partnered with teachers, administrators, district leaders and other researchers to co-design learning progressions as representations of the ways student ideas and engagement in science practices develop over time, and have used these progressions to support teachers' collaborative design of formative assessment tasks. Conducted primarily in high school, this research has explored the ways in which a routine for professional learning, which we have called the *Formative Assessment Design Cycle,* can scaffold the ways teachers navigate student thinking with the support of learning progressions.

Learning progressions

I work from the definition of learning progressions as hypotheses for how student ideas can develop over time, from the lower-anchor, everyday ideas they have when they come to school, to the top-anchor, scientifically accepted understandings they reach after participating in school (Alonzo & Gotwals, 2012). The middle levels chart out various pathways for how these ideas can develop over time. I then take these representations as orienting frameworks for systems of assessment (Shepard et al., 2018); that is, the progressions themselves can help to inform the design of learning sequences, as well as assessment tasks.

We have explored how learning progressions can also be direct supports for teachers as they learn to navigate students' thinking in reform-oriented science teaching, all toward the outcome of broadening participation in science learning. From this perspective, teacher learning, curriculum and assessment design, and students' learning can all be coordinated around and by learning progressions that represent the ideas and practices students are expected to learn within a span of time.

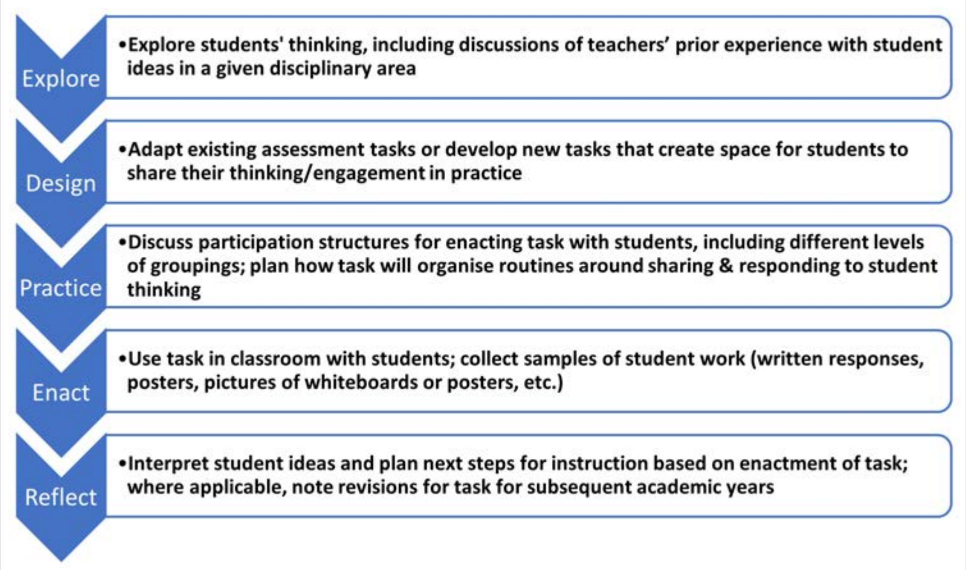
Formative assessment

Formative assessment is a fundamental element of science teaching and learning reforms (NRC, 2014). By creating opportunities for students and teachers to surface and discuss ideas while instruction is in progress (e.g. Black & Wiliam, 1998), formative assessment is an ideal activity to help teachers connect students' ideas and experiences with science, and to adapt and adjust learning experiences to better centre students' interests and needs. The phrase 'formative assessment' itself can refer to both the tasks that teachers use to organise classroom activity, as well as the practices in which students and teachers participate around these tasks, as well as informally in the course of listening and responding to student ideas (Bennett, 2011). Formative assessment tasks are designed to consist of multiple components that create space for students to share their thinking, through open-ended questions, drawing models, writing explanations and other features (Fine & Furtak, 2020; Kang et al., 2014; NRC, 2014). Formative assessment practices are the actions teachers and students engage in to ask each other questions, to push each other's thinking and to better understand each other's ideas (e.g. Cowie et al., 2011). While formative assessment remains a key mechanism for teachers to support student learning, it remains a challenging practice for both pre- and in-service science teachers.

Routines for formative assessment task design: The Formative Assessment Design Cycle

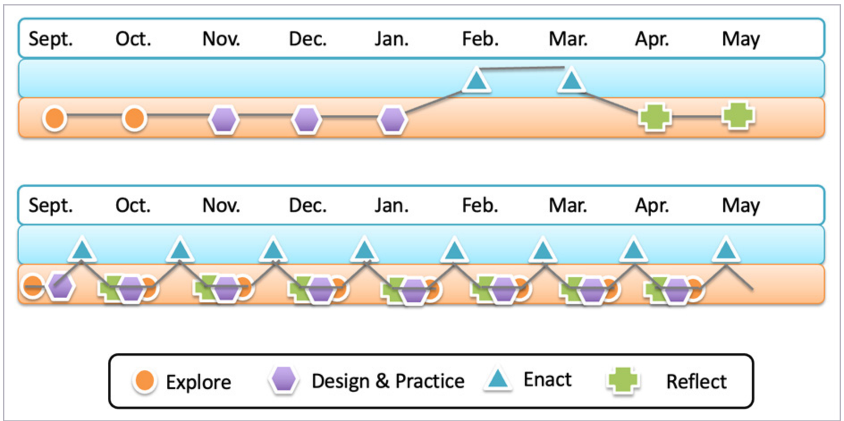
To support teachers' abilities to enact formative assessment, my research team has developed a routine for teachers' collaborative co-design of formative assessment tasks centred on learning progressions. The *Formative Assessment Design Cycle* (Furtak et al., 2014; Furtak & Heredia, 2014) consists of five steps, summarised in Figure 1. First, teachers *explore* students' thinking with the support of a learning progression, and then they *design* formative assessment tasks to embed in units of instruction. Next, they *practice* using those tasks, anticipating how to elicit and respond to student ideas related to the learning progression before *enacting* the tasks with students. Finally, teachers gather to *reflect* upon enactment, using the learning progression to interpret artifacts of student thinking and to plan next steps for instruction.

Figure 1 Formative Assessment Design Cycle



Along with members of my research team and our school district partners, I have used these routines to organise regular meetings of science teacher professional learning communities. As shown in Figure 2, these meetings can be spread out, with an intensive focus on one instructional unit within a school year. Alternatively, meetings can happen more frequently across a nine-month school year, with teachers completing multiple cycles of exploring student ideas, designing and practising using tasks, enacting, and reflecting together.

Figure 2 Frequency of meetings within an academic year



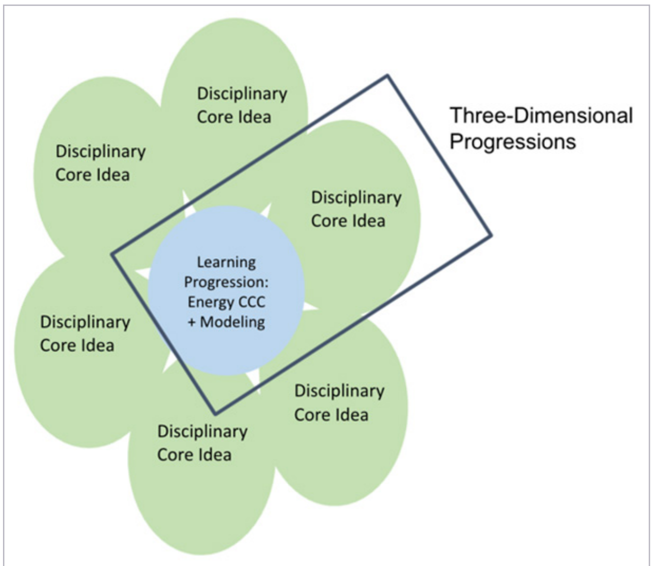
Case example: A learning progression to support vertical alignment in high school science

My team has collaborated with teachers and curriculum leaders in a large culturally, ethnically, linguistically and socioeconomically diverse school district. This collaboration has taken the form of a research-practice partnership (Penuel et al., 2011), in which researchers and practitioners develop mutually beneficial goals to support shifts in classroom practice while also producing knowledge for the broader science education community. The intention of the second study was to examine the ways in which a learning progression could provide opportunities for vertical alignment across high school science courses.

District science leaders were interested in how science teachers could have interdisciplinary conversations about student learning by exploring crosscutting concepts, or concepts and ideas that span across science disciplines (e.g. Nordine & Lee, 2021). One of the crosscutting concepts in the *Next Generation Science Standards* is energy and matter flows, cycles and conservation (NRC, 2012). We developed a new learning progression on the basis of prior progressions created for energy in physics (Neumann et al., 2013) as well as studies of energy as a crosscutting concept (e.g. Park & Liu, 2016). In addition, we integrated the science practice of modelling (Pierson et al., 2017; Schwarz et al., 2009) to create a hypothesised trajectory for how students model energy transfers and transformations across systems (Buell et al., 2019; see Figure 3).

Figure 3 Learning progression for modelling energy

Level	A learning progression for modelling energy flows
5	• Students are able to **generalise their model** to unknown or multiple phenomena, and can **explain limitations of applying the model** to a new phenomenon.
4	• Students **develop a model** that illustrates a mechanism that can explain or predict the phenomenon AND **use** the model to make predictions about how changing one part of the model would influence energy flows elsewhere in the system. • Students can explain how the energy of the system constrains the magnitude of change possible. • Students can **describe limitations of the model** in explaining or predicting the phenomenon.
3	• Students **use** or **develop** a model that relates changes in the phenomenon directly to changes in energy through transfers/transformations by identifying specific indicators. • Students begin to show evidence that their model is accounting for conservation and dissipation. • Model includes energy flows into, within and out of the system.
2	• Students **use** or **develop** a model to illustrate a relationship or pattern between the increase in one form of energy and the decrease in another form, or transferred from one location or object to another. • Students identify the most relevant component and relationships in the model and distinguish between the system and surroundings. • Model focuses on energy flows within the system only.
1	• Students **use** or **develop** a model that shows, through drawings or labels, the components involved in a phenomenon, some (but not necessarily all relevant) energy forms, transfers or transformations.

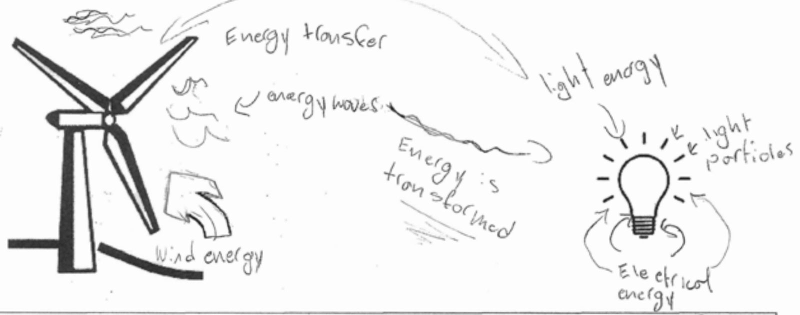
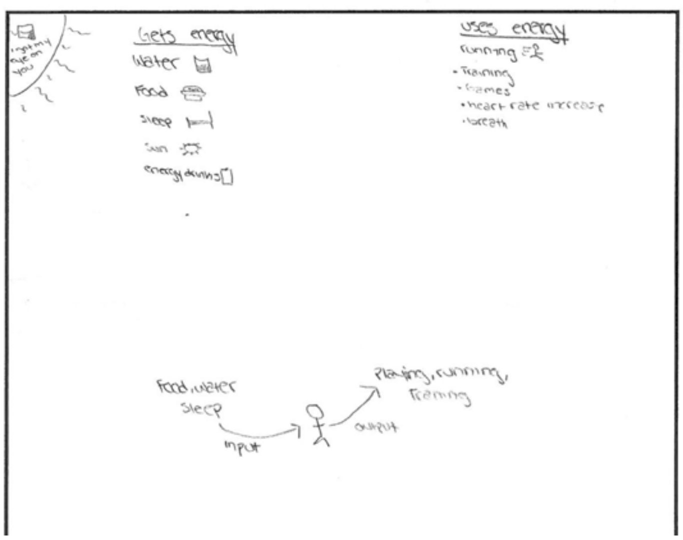


Our intention was that a learning progression focused on an idea that cut across physics, chemistry and biology. It could also create space for students to create models of energy transfers and transformations in different real-life scenarios and opportunities for teachers to support student learning of energy more systematically across grade levels.

We started out by engaging teachers in exploring their own understandings of energy by asking them to model energy transfers and transformations in a simple calorimetry lab. This lab involved teachers creating a simple, inefficient calorimeter and burning a puffed cheese snack, and then drawing models in cross-disciplinary groups. They noticed right away that the language they used to describe energy was different based upon their own content preparation; for example, in chemistry, teachers would talk about thermal energy in exo- and endo-thermic reactions, whereas biology teachers would teach about energy within ecosystems.

On the basis of these shared learning experiences, teachers then moved into discipline-specific groups and used the common learning progression to plan classroom assessment tasks and learning experiences for their own science courses. Teachers developed an array of different formative assessment task types, as shown in Figure 4, to support students' modelling of energy (Buell, 2020). In-depth analyses of teachers' conversations in content-specific teams found that the routines of the formative assessment design cycle facilitated deeper conversations about student thinking across multiple years of the study and supported teachers in designing better assessment tasks to surface student thinking (Henson, 2019). In addition, the tasks facilitated teachers making key connections between students' everyday experiences with energy – such as around exercise and respiration – although this varied across classrooms.

Figure 4 Sample assessment tasks for modelling energy

Physics	**How does energy from the wind power a light bulb?** **Directions: Using the drawing below, add information to show how energy gets from the wind to the light bulb.** ❑ Label the **forms of energy** in this system ❑ Label where energy is **transferred** (moving from one place to another) ❑ Label where energy is **transformed** (changing from one form to another) ❑ Label where energy is **lost** from the system ❑ **Explain** the diagram that you have created  **Explanation** The wind that is created from a wind turbine is transferred into energy waves that convert into electrical energy, that energy is then used to power the source or (other to transform into light energy.
Chemistry	**How do you predict the temperature of the water will change?** With this question in mind, use the provided space below to illustrate the following: • Identify the systems and surroundings by drawing before and after depictions of the lab scenario. • Identify and label the predicted types of energy associated with the lab scenario. • Fill in the LOL diagram on the predicted types of energy, as well as determining a final energy output of the water and the piece of metal at the 10-minute mark of the lab.
Biology	**Part 1** Scenario: The Oakland Raiders are at the Mile High Stadium this weekend. They have had a good season overall, but at this game they aren't playing as well as they usually do. Team members experience increased heart rate, breathing rate and **overall fatigue** (tiredness, low energy). The Broncos win by a large margin. ***Cellular Respiration Equation*** $C_6H_{12}O_6 + O_2 \rightarrow 6CO_2 + 6H_2O + \text{Energy}$ ***Part 1*** Including matter and energy from the cellular respiration equation, draw a model (picture, chart, diagram etc.) to show how the Oakland Raiders running back gets and uses energy to play the game against the Denver Broncos. Show things that you can see, and things that you cannot see. *Use arrows and labels to show the flow of energy and cycling of matter.* 

Take-aways for learning progressions and teacher professional development

As this short case example illustrates, there are multiple opportunities for teachers to collaborate with the support of learning progressions focused on the big ideas and practices of science. While formats for learning progressions vary widely, and their availability is not consistent across science domains, using available learning progressions and adapting them in collaboration with teachers can nevertheless support key conversations about students' opportunities to learn, and to create space to talk together about designing better assessment tasks over time.

This material is based upon work supported, in part, by the National Science Foundation under Grant No. 1561751.

References

Alonzo, A. C., & Gotwals, A. (2012). *Learning Progressions in Science: Current challenges and future directions*. Sense Publishers.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, *5*(1), 7–74. http://dx.doi.org/10.1080/0969595980050102

Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, *18*(1), 5–25. https://doi.org/10.1080/0969594X.2010.513678

Buell, J. Y. (2020). *Designing for relational science practices*. [Doctoral dissertation]. University of Colorado Boulder.

Buell, J. Y., Briggs, D. C., Burkhardt, A., Chattergoon, R., Fine, C., Furtak, E. M., Henson, K., Mahr, B., & Tayne, K. (2019). *A learning progression for modeling energy flows in systems*. Center for Assessment, Design, Research and Evaluation (CADRE).

Cowie, B., Jones, A., & Otrel-Cass, K. (2011). Re-engaging students in science: Issues of assessment, funds of knowledge and sites for learning. *International Journal of Science and Mathematics Education*, *9*(2), 347–366. https://doi.org/10.1007/s10763-010-9229-0

Fine, C., & Furtak, E. M. (2020). A framework for science classroom assessment task design for emergent bilingual learners. *Science Education*, *104*(3), 393–420. https://doi.org/10.1002/sce.21565

Furtak, E. M., & Heredia, S. C. (2014). Exploring the influence of learning progressions in two teacher communities. *Journal of Research in Science Teaching*, *51*(8), 982–1020. https://doi.org/10.1002/tea.21156

Furtak, E. M., Morrison, D., & Kroog, H. (2014). Investigating the link between learning progressions and classroom assessment. *Science Education*, *98*(4), 640–673.

Henson, K. (2019). Designing tools to support teacher activity focused on student thinking to inform instruction. [Doctoral dissertation]. University of Colorado Boulder.

Kang, H., Thompson, J., & Windschitl, M. (2014). Creating opportunities for students to show what they know: The role of scaffolding in assessment tasks. *Science Education*, *98*(4), 674–704. https://doi.org/10.1002/sce.21123

National Academies of Sciences, Engineering, and Medicine. (2015). *Science Teachers' Learning: Enhancing opportunities, creating supportive contexts*. The National Academies Press.

National Research Council. (2012). *A framework for K–12 science education*. National Academies Press.

National Research Council. (2014). *Designing assessments for the Next Generation Science Standards*. National Academies Press.

Neumann, K., Viering, T., Boone, W. J., & Fischer, H. E. (2013). Towards a learning progression of energy. *Journal of Research in Science Teaching*, *50*(2), 162–188.

Nordine, J. & Lee, O. (2021). *Crosscutting concepts: Strengthening science and engineering learning*. NSTA Press.

Organisation for Economic Co-operation and Development. (2017). PISA 2015 Assessment and analytical framework: Science, reading, mathematic, financial literacy and collaborative problem solving, revised edition. PISA, OECD Publishing.

Park, M., & Liu, X. (2016). Assessing understanding of the energy concept in different science disciplines. *Science Education*, *100*(3), 483–516.

Penuel, W. R., Fishman, B. J., Cheng, B., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, *40*(7), 331–337. https://doi.org/10.3102/0013189X11421826

Pierson, A. E., Clark, D. B., & Sherard, M. K. (2017). Learning progressions in context: Tensions and insights from a semester-long middle school modeling curriculum. *Science Education*, *101*(6), 1061–1088.

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654.

Shepard, L. A., Penuel, W. R., & Pellegrino, J. W. (2018). Using learning and motivation theories to coherently link formative assessment, grading practices, and large-scale assessment. *Educational Measurement: Issues and Practice*, *37*(1), 21–34. https://doi.org/10.1111/emip.12189

Identifying and monitoring progress in collaboration skills

Dr Claire Scoular
Australian Council for Educational Research

https://doi.org/10.37517/978-1-74286-638-3_15

Dr Claire Scoular's research interest revolves around the assessment and teaching of 21st-century skills. She has worked with numerous countries to identify and bridge the gap between their assessment, curriculum and pedagogy in relation to skills. In particular, she has led reviews and development of curriculum documentation, developed classroom-based assessment tools and generated teaching resources for skills including problem-solving, collaboration, learning in digital networks, critical thinking, creative thinking and self-regulation. She has been an active lecturer for many years and has developed courses in both traditional formats and online formats, including both the design of the content, and the assessment tasks. Her experience in psychometric assessment, intervention work and research spans across the UK, the Asia-Pacific and South America. She has undertaken consultancies for WorldBank, The Brookings Institution and UNESCO. Dr Scoular has been a Keynote Speaker on multiple occasions, and recently served as a presenter and mentor to delegates at the Y20 Youth Summit.

Abstract

The nature of skills such as collaboration is complex, particularly given that there are internal processes at play. Inferences need to be made to interpret explicit behaviours observed from intentionally designed assessment tasks. This paper centres on the approach to develop hypotheses of skill development into validated learning progressions using assessment data. Understanding a skill from a growth perspective is essential for the effective teaching and development of the skill. The application of Item Response Theory (IRT) allows the interpretation of assessment data as levels of proficiency that we can use to map or monitor progress in collaborative skills.

Collaboration as an important skill for learning

There is increasing demand to work well with others and to work globally (O'Neil et al., 2004). Consequently, collaboration skills that allow effective working in groups have been identified as increasingly important for success in education and work environments (Singh-Gupta & Troutt-Ervin, 1996). As a result, educational research on collaboration has been in abundance in recent years (Griffin & Care, 2015; Organisation for Economic Co-operation and Development [OECD], 2017 Scoular & Care, 2019; von Davier & Halpin, 2013). The OECD's decision to assess collaborative problem-solving (CPS) in the Programme for International Student Assessment (PISA) in 2015 has been a major driver in highlighting the importance of understanding and measuring this skill.

Collaboration has been shown to enhance cognitive development (Webb, 1989; Zhang, 1998) and has been demonstrated to have advantages in encouraging accountability, ability to ask questions and justify responses, flexibility in problem-solving and reflective skills (Baghaei et al., 2007; Soller 2001; Webb et al., 1998). Several prominent researchers highlighted the learning benefits to the individual of interaction with other humans, suggesting that placing learners in a social context is a core strategy for developing complex cognitive skills such as problem-solving competency (Glaser, 1992; Vygotsky, 1986; Wittrock, 1989). When learners work collaboratively to solve problems, they think through the problem and the processes more explicitly during their interaction with others, which leads to a greater conceptual understanding and more effective task management (Darling-Hammond, 2003).

There is research to suggest that learners process information differently when they work in groups compared to working independently (King et al., 1997). Social interactions make learners' understanding explicit and learners usually improve their comprehension through discussion with others, elaborating, and negotiating with others to reach shared understanding (Van Boxtel et al., 2000). Collaborative actions, such as asking questions, peer mentoring, and providing feedback, can help learners to solve problems or finish tasks they may have otherwise not been able to solve or complete and, therefore, allow them to move towards higher levels of proficiency (King et al., 1997). Social interactions while working through complex tasks can provide additional ideas and shared meaning that an individual would not achieve without communicating with others (OECD, 2013).

In education systems around the world, teachers are being tasked with monitoring and improving students' collaboration skills (Scoular et al., 2020). One of the major challenges in that endeavour is identifying exactly what collaboration looks like in the classroom and how proficiency in it can be described. Beyond identifying the importance of collaboration, there is little guidance on where, when, and how to develop, teach and train such skills. This is partly due to a lack of understanding of the nature of the skills, including how it develops and changes over time, and which aspects are fundamental building blocks.

One challenge in measuring skills such as collaboration is that they are complex, particularly given there are internal processes at play. Evidence of ability in such skills is likely to be covert, not directly observable and, therefore, inferences about student ability need to be drawn from demonstrated behaviours observed from intentionally designed assessment tasks. Education systems frequently emphasise summative assessments that centre on overall score points. These serve a purpose but typically do not contribute to individual learning or development of growth in skills. Measuring an innovative domain requires innovative measures. Assessments of progress are an alternative to judging success only in terms of year-level standards. Identifying specifically what a learner needs to know or be able to do can better inform how instructional support can be adapted to support people to progress towards their goals. In other words, the focus should be on learning progressions, rather than just on the scores.

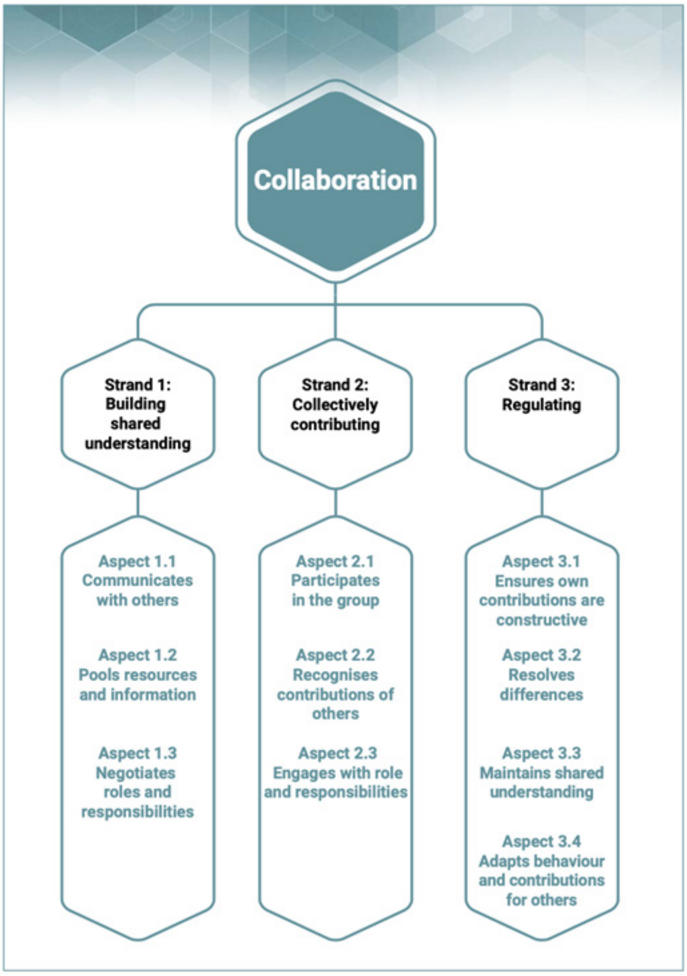
Moving from conceptualisation to assessment data

The definition of collaboration is much more complex than simply working with others. The literature has shifted from a simple definition of collaboration as working in groups, to defining collaboration as an action where two or more learners pool knowledge, resources and expertise from different sources in order to reach a common goal. The distinction between interdependence and independence provides some insight into the nature of collaboration. While the focus of team or group work literature has been on independent teams where learners work in relative isolation, interdependent teams rely on the actions of others and cannot perform activities independently (von Davier & Halpin, 2013). Collaboration is interdependent. There is shared responsibility and an active division of labour. Although there are different definitions of collaboration presented in the literature, similar components can be identified in each (Hesse et al., 2015; OECD, 2017). For example, due to the nature of collaboration, the participation of each learner and their level of engagement with an activity directly impacts on the effectiveness of the collaborative group as a whole.

The definition of a skill has implications for assessment task design, and good measurement practice indicates that a clear definition of the domain the assessment is measuring should be identified before task design begins (Scoular et al., 2017). As a means to support teachers in understanding the skill in a detailed way, the Australian Council for Educational Research (ACER) produced a skill development framework for collaboration (Scoular et al., 2020). The purpose of the ACER framework is to establish a common terminology for describing collaboration, taking into consideration existing assessments of collaboration, and providing a structure that is suited for the ongoing assessment and teaching of collaboration. The framework breaks collaboration down into strands that are then further elaborated as aspects (see Figure 1). Within this collaboration definition, there is specification of three strands: building shared understanding, collectively contributing, and regulating, and within each strand there are associated aspects.

The framework also includes skill development levels that outline how growth in a particular aspect can be demonstrated, and how learners move from early to more advanced application and understandings. These levels support efforts to measure attainment and monitor learner growth over time, and are underpinned by an understanding that learners of the same age and in the same year of school can be at very different points in their learning and development. For this reason, the levels are deliberately not linked to years of schooling. The levels were initially hypothesised, using literature, research and expert judgement to build an understanding. To ensure an evidence-based approach, data from a number of assessments were then compared and contrasted to the hypothesised levels. Data were drawn from a number of sources, including ACER assessments and the PISA-CPS 2015 assessment data. The remainder of this paper will demonstrate an example of the approach taken.

Figure 1 Collaboration definitional framework



Source: Scoular et al (2020)

Wright map

Item Response Theory (IRT) is a psychometric analysis often applied to assessments to examine item quality and identify student ability. One particular IRT output, the Wright map (Wilson & Draney, 2002), is helpful for visualising the interaction between different items, and their estimated difficulties, in relation to student ability. The Wright map places coded responses to items and learner estimates onto a single scale, using logits as the scaling unit (an arbitrary unit used to enable location of the two variables on the same metric). It presents the items in increasing order of sophistication in relation to learner ability and can be viewed as two vertical histograms.

An example Wright map for a collaboration assessment included in our validation process is presented in Figure 2. The left side of the figure displays the distribution of learner ability as a histogram and the right side of the figure displays the distribution of the item difficulties. Items 32.2, 38.2 and 50.2 are at the top of the map indicating they are the most difficult. Items 11.1, 12, 19.1 and 40 are at the bottom of the map indicating they were the easiest. The learner ability distribution extends slightly lower than these items so it is difficult to discriminate between those learners at this very low level, although there are only a few of them.

Figure 2 Wright map for a collaboration assessment

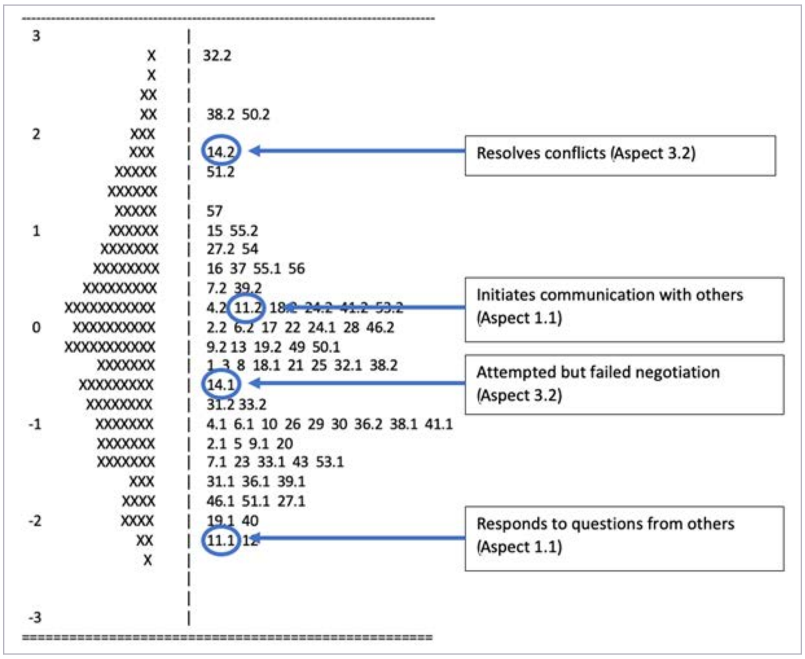
```
--------------------------------------------------------------------------------
 3                  |
              X     | 32.2
              X     |
             XX     |
             XX     |  38.2 50.2
 2          XXX     |
            XXX     |  14.2
          XXXXX     |  51.2
         XXXXXX     |
          XXXXX     |  57
 1       XXXXXX     |  15 55.2
        XXXXXXX     |  27.2 54
       XXXXXXXX     |  16 37 55.1 56
      XXXXXXXXX     |  7.2 39.2
    XXXXXXXXXXX     |  4.2 11.2 18.2 24.2 41.2 53.2
 0   XXXXXXXXXX     |  2.2 6.2 17 22 24.1 28 46.2
    XXXXXXXXXXX     |  9.2 13 19.2 49 50.1
        XXXXXXX     |  1 3 8 18.1 21 25 32.1 38.2
      XXXXXXXXX     |  14.1
       XXXXXXXX     |  31.2 33.2
-1      XXXXXXX     |  4.1 6.1 10 26 29 30 36.2 38.1 41.1
        XXXXXXX     |  2.1 5 9.1 20
        XXXXXXX     |  7.1 23 33.1 43 53.1
            XXX     |  31.1 36.1 39.1
           XXXX     |  46.1 51.1 27.1
-2         XXXX     |  19.1 40
             XX     |  11.1 12
              X     |
                    |
                    |
-3                  |
================================================================================
```

Moving from assessment data to understanding skill development

The Wright map gives a numerical scale of increasing proficiency or ability in the domain, and it supports a qualitative understanding of different parts of that scale by looking at the differently coded responses to the items, or statements about the substance of differently coded responses to items, that are ranged along the scale. When a learner's ability is estimated, the probability of item response success can be identified by referencing the corresponding learner and item locations on the Wright map. This unidimensionality, and the relative locations of items and learners along its continuum, provides crucial information in understanding the skill and its structure. That is, the data presented by the Wright map allow us to understand the domain as a continuum.

As presented in Figure 3, for example, the location of the coded responses to the items and what aspects they are mapped to can tell us a great deal about proficiency in collaboration. Item 11 is mapped to Aspect 1.1 ‘communicates with others’ in the collaboration framework. A coded response of 1 on this item (denoted 11.1) is specifically the behaviour ‘responding to questions from others’ and is considered within the ACER framework to be a behaviour associated with relatively low proficiency in collaboration. This item is quite far down the Wright map suggesting that it was a fairly easy item, and that corroborates our assumption. By comparison, a coded response of 2 (denoted 11.2) is a behaviour associated with relatively higher proficiency within this aspect, ‘initiates communication with others’, and it is much higher up the Wright map suggesting it is indeed a more difficult item. This suggests that there are different levels of proficiency within aspects; for example, less proficient learners would only communicate with others when they are responding to questions, whereas more proficient learners would initiate communication with and between group members. Similarly, difference levels of proficiency can be interpreted from item 14, which is mapped to Aspect 3.2 ‘resolves differences’. A lower score on this item (denoted 14.1) is in the lower half of the Wright map and is measuring learners attempting but failing negotiations. By comparison, a higher score on this item (denoted 14.2) is one of the most difficult items at the top of the Wright map and is measuring ‘resolves conflicts’. These item locations suggest that less proficient learners attempt to negotiate with others but often cannot resolve differences, in contrast to highly proficient learners who can resolve differences.

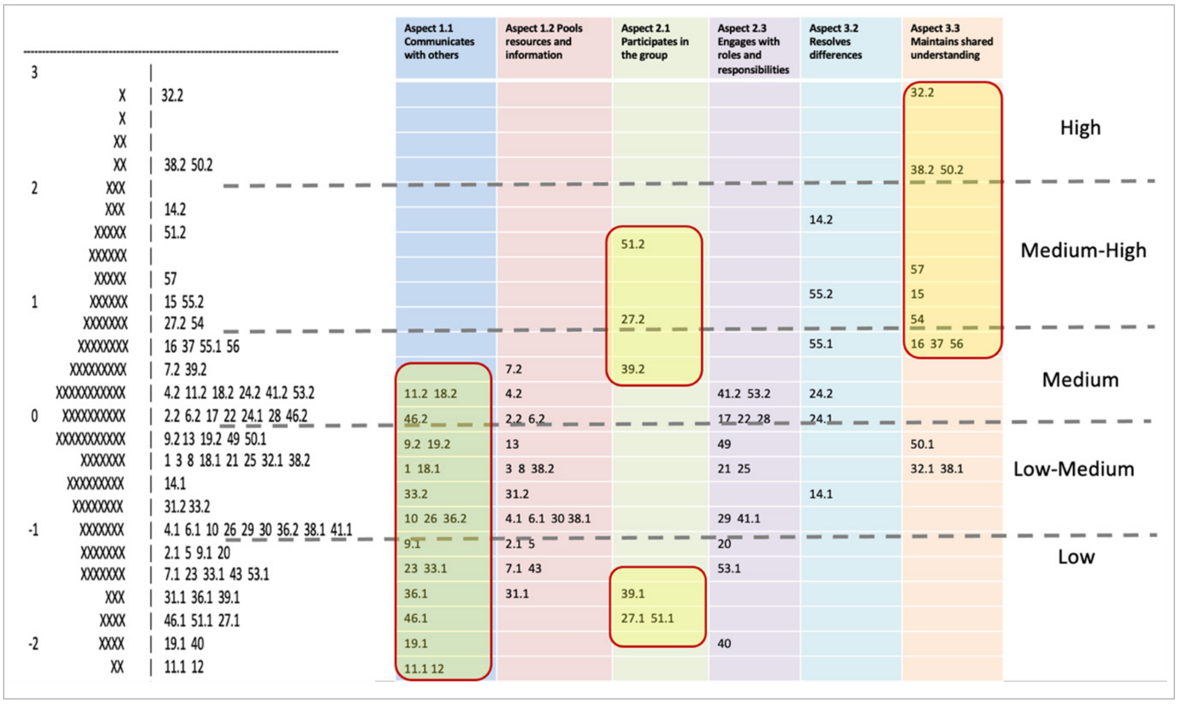
Figure 3 Interpreting specific items’ difficulty in the Wright map



The Wright map presented in this paper allows interpretation of the ordering of the items as a progression of collaborative skill. The positioning of items on that continuum can be interpreted to better understand the domain at different levels of proficiency. This allows meaningful information about what learners can and cannot do given their ability estimates to be relayed back to educators, and to learners themselves. As the item examples in this paper relate to behaviours, it can provide educators with real-time data regarding the social and cognitive behaviours their learners are demonstrating. This level of information extends beyond that gathered and interpreted through static tasks or multiple-choice tests. If the expectation of educators is to teach skills such as collaboration, then detailed data about what their learners can demonstrate is imperative and can complement the evidence already being gathered. Having knowledge of the behaviours expected at higher levels of ability can help educators to develop and implement instructional activities that are targeted to learner needs.

It is also possible to interpret the Wright map as levels of proficiency by interpreting groups of items where they appeared to cluster at similar levels on the scale. Figure 4 identifies cut off points on the scale where natural breaks appear between clusters of items that have similar item difficulty levels. These lines indicate where a discernible change in item difficulty was associated with a change in the kind of skill (or ability) required to demonstrate the associated behaviours. These cut scores and clusters of indicators enable an interpretation of the scale as a developmental progression (Griffin, 2007).

Figure 4 Interpreting proficiency from the Wright map



Interpretation of the items and where they are clustered on the map can lead to the production of text descriptions of what is occurring at each level. The descriptions can be interpreted as level of progress in collaboration from novices (A) to experts (E). For example, novices would work independently, unaware of the benefit of engaging with peers, and thus would not be likely to solve a problem collaboratively. As learners increase in proficiency, they engage more with others and gradually realise the benefit of collaboration. Expert collaborators depend upon their peers, develop a mutual understanding of problems or tasks, and work strategically through them together. Negotiation is a critical component of collaboration, but only proficient learners can harness conflicts towards a positive outcome.

A broader sense of the skill

It is important to note that the skill just discussed is interpreted through the lens of a single assessment. It is likely that not every item in a cluster will tell the same conceptual story as the others. While the item may fit statistically, it can be omitted from this part of the interpretation on conceptual grounds. Similarly, no single assessment is expected to measure the entirety of a domain. For example, the assessment example used in this paper did not measure Aspect 3.4 'adapts behaviours and contributions for others' or Aspect 1.3 'negotiates roles and responsibilities'. Therefore, interpreting a single assessment in terms of proficiency levels will likely present gaps in the representation of the skill as a whole.

In order to get a broader sense of a skill, learning progressions can be developed by drawing on response data from multiple assessments and expert judgement. Learning progressions are not the same as, or extensions of, learning standards; rather, they focus explicitly on the building blocks learners need to master before they can achieve complex skill forms. When assessments provide information about where learners are in their understanding at the time of assessment, they also provide a basis for monitoring individual progress over time. However, it is important to note that learning progressions are not intended to describe a single pathway to achieving proficiency in a skill that it is assumed all learners will follow. Instead, a learning progression describes a typical pathway that provides a good starting point for deeper interrogation of learners' unique pathways.

Table 1 presents an excerpt from the ACER skill development levels for the three strands of collaboration. This representation of learning progress for collaboration has been developed and validated using data from multiple assessments, including the Wright map presented in this paper (Scoular et al., 2020). The levels are intended to support understanding of the skills and the ways in which they develop. They can also support teachers to identify gaps in a learning area, where some learners may require further assistance in order to move learning forward along this path.

Table 1 Excerpt from collaboration learning progression

Skill level	Building shared understanding	Collectively contributing	Regulating
Medium	Learners ask for justification of responses or perspective provided (Aspect 1.1)	Learners acknowledge that others may have a different perspective, and that based on these perspectives, others' contributions may be beneficial to the group as a whole. They understand and incorporate the contributions of others into their own work. (Aspect 2.2)	Learners identify their own strengths and weaknesses in relation to the progress of the group task as a whole. (Aspect 3.1) Learners make constructive but unsuccessful attempts to resolve differences. (Aspect 3.2) Learners act to maintain shared understandings, such as by reiterating and finalising goals, strategy, and roles in more complex tasks. (Aspect 3.3) Learners require feedback from others or explicit requests before they modify or tailor their communication style or behaviour. (Aspect 3.4)
Low–mid	Learners ask questions or for clarification from others. They will communicate about the related task and respond to contributions of others. (Aspect 1.1) Learners identify that they may not have all of the information required and pool some resources and information with others. (Aspect 1.2) Learners negotiate roles but without considering the expertise, information or skills help by other group members. (Aspect 1.3)	Learners participate in all necessary tasks throughout the tasks. Learners maintain a single strategy throughout. Learners collaborate successfully to achieve a straightforward goal. (Aspect 2.1) Learners understand that others may have an alternative perspective. They listen to and acknowledge the perspective of others. (Aspect 2.2) Learners show a willingness and readiness to be involved in the group. They take responsibility for some of the actions determined by their role and provide feedback on their individual task. (Aspect 2.3)	Learners reflect on the quality and relevance of their own contributions. (Aspect 3.1) Learners discuss differences of opinion or perspective with others and give careful consideration of the views of others. They comment on differences but are often unable to resolve them. (Aspect 3.2) Learners act to maintain shared understanding through reiterating goals, strategy and roles in basic tasks. (Aspect 3.3)

Source: (from Scoular et al., 2020)

Conclusion

It is increasingly apparent that our understanding of complex skills needs to be enhanced in order to meet the demand of 21st-century education (Griffin & Care, 2015; OECD, 2017, Scoular & Care, 2019; von Davier & Halpin, 2013). Complex skills such as collaboration can be difficult to teach and learn, but the approach presented in this paper demonstrates that robust measurements can be developed that provide insight into how these skills can be demonstrated. Further, assessment data can actually provide more information and improve understanding of such complex skills. The ACER skill development framework for collaboration sets out behaviours and processes that can be associated with aspects of collaboration, and assessment data provide validation of this. No single assessment can paint the larger picture, but each individual assessment of collaboration can provide a piece of the puzzle in understanding this complex skill. Interpretations of the data visualisations such as the Wright maps can indicate how different proficiencies of collaboration might be demonstrated. Assessment of such skills, particularly in relation to growth, can shed light on how to appropriately situate teaching interventions and to identify learning in an innovative domain. The work ahead is iterative. Our understanding of 21st-century skills needs to be documented so that tasks can be designed appropriately to match this understanding. The more we develop robust assessments of 21st-century skills, the more we can corroborate, validate, and evolve that understanding. If we are to continue to value these skills in the 21st century and beyond, efforts in understanding how they develop and what different levels of proficiency look like will need to be applied.

References

Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting collaborative learning and problem solving in a constraint-based CSCL environment for UML class diagrams. *International Journal on Computer-Supported Collaborative Learning*, *2*(2), 159–190. https://doi.org/10.1007/s11412-007-9018-0

Darling-Hammond, L. (2003). *The learning classroom: Theory into practice*. School of Education, Standford University. https://www.learner.org/courses/learningclassroom/support_pages/index.html

Glaser, B. (1992). *Basics of grounded theory analysis*. Sociology Press.

Griffin, P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation*, 33, 87–99. https://doi.org/10.1016/j.stueduc.2007.01.007

Griffin, P., & Care, E. (2015). *Assessment and teaching 21st century skills: Methods and approach*. Springer. https://doi.org/10.1007/978-94-017-9395-7

Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills: Methods and approach* (pp. 37–56). Springer. https://doi.org/10.1007/978-94-017-9395-7_2

King, F. J, Rohani, F., & Goodson, L. (1997). *Statewide assessment of listening and verbal communication skills, information literacy skills, and problem-solving skills*. Florida State University.

O'Neil, H. F., Chuang, S., & Chung, G. K. W. K. (2004). Issues in the computer-based assessment of collaborative problem solving. *Assessment in Education*, *10*, 361–373. https://doi.org/10.1080/0969594032000148190

Organisation for Economic Co-operation and Development. (2013). *PISA 2015: Draft collaborative problem solving framework*. http://www.oecd.org/callsfortenders/Annex%20ID_PISA%202015%20Collaborative%20Problem%20Solving%20Framework%20.pdf

Organisation for Economic Co-operation and Development (OECD). (2017). *PISA 2015 Results (Volume V): Collaborative problem solving*. OECD Publishing. https://doi.org/10.1787/9789264285521-en

Scoular, C., & Care, E. (2019). A generalized scoring process to measure collaborative problem solving in online environments. *Educational Assessment*, 1–22. https://doi.org/10.1080/10627197.2019.1615372

Scoular, C., Care, E., & Hesse, F. (2017). Designs for operationalizing collaborative problem solving for automated assessment. *Journal of Educational Measurement*, *54*(1), 12–35. https://doi.org/10.1111/jedm.12130

Scoular, C., Duckworth, D., Heard, J., & Ramalingam, D. (2020). *Collaboration: Skill development framework*. Australian Council for Educational Research. https://research.acer.edu.au/ar_misc/42

Scoular, C., Eleftheriadou, S., Ramalingam, D., & Cloney, D. (2020). Comparative analysis of student performance in collaborative problem solving: What does it tell us? *Australian Journal of Education*, *64*(3), 282–303. https://doi.org/10.1177/0004944120957390

Singh-Gupta, V., & Troutt-Ervin, E. (1996). Preparing students for teamwork through collaborative writing and peer review techniques. *Teaching English in the Two Year College*, *23*(2), 127–136.

Soller, A. (2001). Supporting social interaction in an intelligent collaborative learning system. *International Journal of Artificial Intelligence in Education*, *12*(1), 40–62.

Van Boxtel, C., Van der Linden, J., & Kanselaar, G. (2000). Collaborative learning tasks and the elaboration of conceptual knowledge. *Learning and Instruction*, *10*(4), 311–330.

Vygotsky, L. S. (1986). *Thought and language*. The MIT Press.

Webb, N. M. (1989). Peer interaction and learning in small groups. *International Journal of Education Research*, *13*, 21–39. https://doi.org/10.1016/0883-0355(89)90014-1

Webb, N. M., Nemer, K. M., Chizhik, A. W., & Sugrue, B. (1998). Equity issues in collaborative group assessment: Group composition and performance. *American Educational Research Journal*, *35*(4), 607–651. https://doi.org/10.3102/00028312035004607

Wilson, M., & Draney, K. (2002). A technique for setting standards and maintaining them over time. In S. Nishisato, Y. Baba, H. Bozdogan, & K. Kanefugi (Eds.), *Measurement and multivariate analysis* (pp. 325–332). Springer-Verlag.

Wittrock, M. C. (1989). Generative processes of comprehension. *Educational Psychologist*, *24*(4), 345–376. https://doi.org/10.1207/s15326985ep2404_2

Zhang, J. (1998). A distributed representation approach to group problem solving. *Journal of American Society of Information Science*, *49*(9), 801–809. https://doi.org/10.1002/(SICI)1097-4571(199807)49:9<801::AID-ASI5>3.0.CO;2-Q

Reporting student progress: What might it look like?

Hilary Hollingsworth
Australian Council for Educational Research

Jonathan Heard
Australian Council for Educational Research

Anthony Hockey
Principal, St Mary's Primary School

Tegan Knuckey
Assistant Principal, Balwyn High School

https://doi.org/10.37517/978-1-74286-638-3_16

Hilary Hollingsworth is a Principal Research Fellow in the Educational Monitoring and Research Division at ACER. She has 30 years of experience working in a wide range of national and international educational contexts including schools, universities, research organisations, government education departments and private education service organisations. Her expertise is in teaching and learning, teacher education and professional development, classroom observation frameworks and the use of video, teacher feedback, teaching quality, assessing student learning and communicating student progress. Hilary is a member of ACER's Centre for Assessment Reform and Innovation (CARI) and led the Communicating Student Learning Progress *project.*

Jonathan Heard is also a Research Fellow in the Educational Monitoring and Research Division at ACER. He has 15 years' experience as a high school teacher, holding leading teacher positions in institutional coaching and pedagogy. His current work foci at ACER include the assessment of general capabilities and critical thinking, and researching teaching and learning initiatives in school education. Jonathan is also a member of CARI and was a key researcher in the Communicating Student Learning Progress *project.*

Abstract

The Communicating Student Learning Progress review produced by ACER in 2019 set out recommendations for schools and systems to improve the way schools report on student learning, in particular learning progress. Two case study schools from Victoria – a Catholic primary school and government secondary school – discuss changes they've made to their student reporting processes, in response to the review's recommendations. Further research is recommended into how schools are rethinking reporting to engage students and parents in monitoring learning growth.

Introduction

In 2019, a review of student reporting in Australian schools was published as the culmination of a research project undertaken by a small team funded internally through the Australian Council for Educational Research (ACER) Centre for Assessment Reform and Innovation (CARI). Titled *Communicating Student Learning Progress* (Hollingsworth et al., 2019), the review sought to understand the extent to which reporting practices in Australian schools met the stated aim of communicating both achievement and progress (Australian Education Regulation, 2013), with 'progress' understood as a student's gain, growth or increasing proficiency along a continuum of learning as measured over time.

One outcome of the project was a set of eight recommendations to schools and systems about how student reporting might more successfully achieve the goal of communicating the progress students make in their learning, as well as their performance in subjects and assessments. An abridged version of these recommendations is contained in Figure 1.

Figure 1 Recommendations for communicating student learning

1. Schools and systems should use consistent terminology to communicate about student learning.
2. Student reporting should be continuous and aligned to the assessment cycle.
3. Student reporting should explicitly represent and communicate learning progress.
4. Student reporting should explicitly communicate student learning against expectations.
5. Student reporting should clearly articulate how performance ratings are defined.
6. Student reporting should present information that is accessible and provides different levels of detail.
7. Student reporting should include specific directions for future learning.
8. Methods used to communicate student learning should have distinct but complementary purposes.

In addition to these recommendations to schools, the project team proposed recommendations for future research, including to identify and investigate work being undertaken in schools on the design and use of reporting formats that better support, monitor and represent student progress. The aim of such research would be to determine what is possible to achieve and what 'works', for teachers, parents and students.

Nearly two years on from the publication of the report, several questions relevant to this potential research direction have emerged:

- How are schools making use of these recommendations?
- What changes have schools made to their reporting processes?
- What have been the effects of these changes and what have been the challenges?

This paper seeks to explore these questions by presenting the recent work of two schools in Melbourne – St Mary's Primary School, Williamstown and Balwyn High School – and their efforts to reform aspects of their student reporting systems.

Both schools are focusing on the ultimate goal of more effectively communicating learning progress in reports, with the view to making reporting more meaningful for students and parents. To this end, they are both aiming to address recommendation 3 outlined in Figure 1. However, they were asked to describe which of the other recommendations they have focused their attention on in their efforts to make improvements, and what the results have been thus far.

Towards making more effective comments related to progress

– Anthony Hockey, Principal, St Mary's Primary School, Williamstown

Which of the eight recommendations from the Communicating Student Learning Progress report were you most interested to address? Why?

It is a matter of taking small steps. Our main objective was to work towards making our report comments more effective, so our initial focus was on recommendations 1, 3 and 4.

What we found was that our school reports were very traditional in that they tended to report on student achievements. The reports listed things the student had been working on and provided a sentence at the start or finish of the comment that let parents know how well the student had achieved in the area. For example, 'Student A has a good understanding of place value to two decimal places'. What the parents did not know from this is what the student knew beforehand, so was this *good* progress, below the expected level of growth, or was it in fact excellent progress?

Quite often the parent did not fully understand the reference to the curriculum standard or the teacher terminology. Feedback from the parents suggested there was too much 'teacher talk' and that they did not really understand what was happening with their child after reading the comments. Parents were also telling us that our comments were way too long.

What we wanted was effective comments that told parents about the progress their child was making.

What changes did you make?

We made two significant changes. First, we created the opportunity for more student voice in the school reports. Students in Years 3 to 6 now input directly into their school report. Students give themselves a mark using the same rubrics as the teachers and their self-assessment marks appear next to the teacher's. This year we have expanded the scope of student input in their reports: the students are also going to write their own comment that will appear next to the teacher's.

Second, we changed the structure of how the comments are written. In our Literacy and Numeracy comments, we found we were listing a whole lot of what we had taught and used a lot of teacher jargon, but parents were not learning what they really wanted to know. We have simplified the comment to answer questions like, what is something the student did well in, what is something they are still struggling with, what steps do they need to take to improve? Our aim is for our comments to become shorter but more effective.

What changes have you seen as a result?

Some of our changes have been obvious - our students self-assess in grades 3 to 6, and in the upcoming reports they will self-comment. On the whole the teacher comments are shorter, however some teachers still want to write lengthy comments.

The biggest change relates to simplifying how comments are written. Some staff have really taken on the idea that they are telling the journey of a student's learning rather than the end point. Their comments talk more of what they can do now, compared to what they were doing previously, so the parent can see the progress made. For some teachers it was simply adding the word 'now', that changed their thinking. What can the student do now that they previously could not?

Our work in this area also led us to include students in the parent–teacher interview, something that had not been done in the school previously. This has meant that the students are gaining greater understanding of their own learning needs.

What challenges did you encounter, or what is still a challenge?

The challenge we have come across is shifting the philosophy of staff members who have written reports the same way for many years. We would not say our comments are perfect and we are still having teachers struggling to change how they have written reports. Our teachers who have been teaching for longer are the ones finding the change the hardest. Their fear appears to be breaking away from listing what they have been teaching, it is as if they feel they need to write all this to prove to the parent community that they have taught something. Comments like 'Our parents expect to see a long comment', or 'Our parents will not be happy' come from the ones who are struggling the most to change.

Towards presenting information that is accessible and focused on future learning

– Tegan Knuckey, Assistant Principal, Balwyn High School

Which of the eight recommendations from the Communicating Student Learning Progress report were you most interested to address? Why?

In our work at Balwyn High School rethinking our reporting practices, we read and considered all the recommendations, however two in particular – recommendations 6 and 7 – emerged as priorities.

In terms of recommendation 6, feedback received from our parents was that information we reported was not always accessible. Parents felt that the comments made by teachers on Common Assessment Tasks (CATs), which were reported as part of our continuous reporting cycle, were too 'task specific' or 'jargony' for them to be able to understand. Beyond Year 7, parents said they did not engage much with comments in continuous reports, and quantitative data from our school management system (Compass) indicated that parents did not frequently access semester reports.

In regard to recommendation 7, whilst the comments associated with a student's performance on CATs indicated how the student could improve, they often pertained specifically to improvement in the task just completed and, therefore, were not necessarily transferable to future tasks in the subject. As a result, parents did not feel that the comments supported them to support their child's learning.

Parents also indicated that they were most interested in how their child presented in class – Did they work hard? Did they ask questions? Did they contribute? Were they organised? There is often a (perceived) correlation between 'learning behaviours' and student learning progress and our parents really wanted to know what their child was like in class, believing this information could support them to help their child improve. The 'effort' and 'behaviour' judgements we were making, by contrast, were considered too superficial to provide much guidance.

To aid us, and students and parents, in understanding what the next steps of future learning might be for a learner, we developed learning continua for each subject. While based upon the Victorian Curriculum achievement standards, these continua more readily suggested what a student might need to work on next in order to develop greater mastery of key skills and knowledge that run through the curriculum right up to VCE.

What changes did you make?

We wanted students to value learning and use every opportunity to demonstrate what they were able to do so they could make progress in their learning. Our previous assessment and reporting system that placed value solely on three large assessments per subject each semester did not enable this. Therefore, we removed CATs from the vernacular and from semester reports, implementing a system that reflected that every lesson every day was an opportunity for students to show what they know and can do. Instead of continuous reporting to parents on CATs, we now complete and distribute 'Reporting Points' twice per semester (and a semester report at the end of each semester).

Reporting Points are similar in appearance to semester reports in that there is one page per subject. On this page is a generic comment regarding what Victorian Curriculum strands/sub-strands have been covered in the period, an individual comment regarding how the students can continue to make learning progress in the subject, and learner profile judgements (based on statements aligned with our school's shared learning norms).

What changes have you seen as a result?

Many students are more engaged in more classes, knowing that everything they do is valued. They are starting to understand their role in the formative assessment processes being used. Students are also less anxious about assessment, knowing they are provided multiple opportunities to demonstrate their knowledge and skills, and that being able to demonstrate these under timed conditions is not the 'be all and end all'.

Parent feedback suggests that the Reporting Point document allows them to see patterns in how their child approaches learning so they can have productive conversations at home about how to move forward.

What challenges did you encounter, or what is still a challenge?

Constructing and trialling the use of the learning continua we developed, amid all of the other day-to-day tasks that leaders and teachers need to engage in, required intentional investment and strong distributed leadership.

Some teachers are still clinging to the concept of CATs and are not providing their students with as much feedback along the learning journey as other teachers.

Teachers' ability to use technology for pedagogy, assessment and reporting has been a strong professional learning focus, with the 2020 period of remote learning being an important catalyst for a lot of the growth in this area. However, there are still large gaps in the competence of teachers in this area.

The reporting platform we use, Compass Continuum, is still a little clunky (though improving) and does not visually represent progress in a way that is obvious or explicit for parents – they need to hover over data points and know how to use the program in order to more fully access the information available.

Conclusion

The perspective of these two schools suggests there is a wider acknowledgement that reporting can be a valuable instrument to engage parents and students more in the learning process. The challenges they describe point to some significant and complex issues associated with reorienting the purpose of reporting away from being a retrospective summary of achievement, and changing associated long-held practices.

It remains an area of research interest to continue to identify schools, systems and providers who are engaged in reporting reform and who are reimagining what communicating to parents about their child's learning might look like; to capture their practice and identify what works for whom and why.

Anthony Hockey is the Principal of St Mary's Primary School in Williamstown, Victoria.
Tegan Knuckey is Assistant Principal at Balwyn High School in Victoria.

References

Australian Education Regulation. (2013). https://www.legislation.gov.au/Details/F2019C00086

Hollingsworth, H., Heard, J., & Weldon, P. (2019). Communicating student learning progress: A review of student reporting in Australia. ACER.

How might we identify and measure learning progression in history?

Dr Louise Zarmati
University of Tasmania

https://doi.org/10.37517/978-1-74286-638-3_19

Dr Louise Zarmati is a Lecturer in Humanities and Social Sciences Education at the University of Tasmania in Australia. She has had a varied career as a teacher, archaeologist and museum educator. Louise's research focuses on history pedagogy in primary and secondary classrooms and museums, as well as the integration of archaeology and Indigenous knowledges into the school curriculum.

In 2009–10 Louise worked for ACARA as a writer of the Australian Curriculum: History (K-10) and Senior Ancient History Courses. In 2018 she was invited by the OECD to conduct research on international examples of learning progression in History.

Introduction

As recently as the 1990s, history was taught in many schools and universities in Western countries as 'content', or what is now called 'historical knowledge'. History was presented as a succession of dates and events that students were expected to memorise and regurgitate in formal examinations. This narrow focus on content made history tedious for many young people and it was not unusual to hear them complain that they did not like the subject because they could not remember a string of boring dates and facts.

Of course, factual recall is an important part of understanding history as well as other subjects, but if factual recall is considered the main purpose of learning history, then memorisation becomes the key skill learned, not a broader understanding of the nature of history itself.

Fortunately, the study of history has changed dramatically since the 1990s, from a simple focus on the memorisation of people, dates and events to the cognitive understanding of historical concepts and the use of historical skills.

What is 'history'?

History is a dynamic subject that is constantly reshaped in response to changes in research and interpretation over time. Postmodernists remind us that history is a construct that is subject to the variables of time and place. Historical knowledge is geographically and temporally specific and can differ within nations, states and local regions, depending on which knowledge (such as people, dates, and events) is considered at the time to be historically significant.

The processes of 'doing' history

In the 1970s and 1980s, British researchers undertook a study that examined how students aged 7 to 14 years understood history (Lee, 2005; Lee & Ashby, 2000; Lee et al., 1996). They separated the learning of history into two distinct areas:

1. **Knowing about history – Substantive knowledge and understanding**

   This encompasses knowing the details of history, such as names, dates, people, events and places. It also includes the understanding of broad concepts such as colonisation, imperialism, peasant, revolution, ideology, kingship, society, liberty, and feudalism.

2. **Doing history – Procedural knowledge**

   This encompasses the concepts and skills you need to be able to do history. These include understanding concepts such as evidence, continuity and change, and cause and effect in order to make sense of the past. It also requires the ability to describe or explain concepts, and to analyse and evaluate the usefulness and reliability of sources.

Substantive and procedural knowledge build on each other and should function together. It is impossible for students to understand or make use of procedural knowledge if they have no knowledge of the substance of the past.

Historical inquiry

Students use historical knowledge and skills to investigate the past through the process of historical inquiry. It is the heuristic used by educators to teach history. History is problematised as a dilemma, conflict, mystery or contradiction that is analysed, dissected and interpreted.

During the process of historical inquiry, students can do the same work as academic and professional historians but without the same sophistication and complexity. They can ask historical questions, identify contradictions and conflicts, and develop interpretations supported by historical evidence.

Fundamental to historical inquiry is the interrogation and critical evaluation of primary sources, which can be written, visual or archaeological. Students analyse written primary sources such as eyewitness accounts, diaries and newspaper reports, and images such as photographs, postcards, and paintings. They can also examine artefacts such as pottery, weapons, statues, coins, and jewellery, and old or ancient objects of everyday life.

Students also learn to critically evaluate secondary sources, which are sources created after the historical event by someone who did not participate in or experience the event first-hand. Scholarly books, textbooks, research articles and documentaries are examples of secondary sources.

What does it mean to 'think historically'?

Students think historically when they use primary sources as evidence about what happened in the past. They demonstrate their ability to understand different interpretations of the past and, ultimately, use historical evidence to develop their own interpretations.

In order to understand progression in history, students should use a combination of substantive and procedural knowledge in the process of historical thinking. It is a student's ability to demonstrate that they can think historically that is measured as evidence of learning progression in history.

What is learning progression in history?

Learning progression is a continuum that measures advances in learning by tracking development from early learning to more sophisticated levels of mastery. Mathematics relies on an understanding of empirical knowledge and concepts in a hierarchical sequence; students need to understand (or master) one mathematical concept before they can proceed to the next.

In comparison, progress of understanding in history is not necessarily hierarchical because it is based on mastery of concepts and skills rather than historical knowledge, which is geographically and temporally variable.

With history, it is not necessary to progress sequentially from one concept in order to comprehend another; learning is measured by mastery of levels of complexity within each skill or understanding and mastery can be concurrent and interrelated.

The essential characteristic of progression in history is that students can demonstrate an increase in their cognitive ability to think analytically and critically.

Learning progression in the Australian Curriculum: History

Substantive and procedural knowledge are differentiated in the Australian Curriculum as:

- knowledge and understanding (substantive)
- inquiry and skills (procedural)

The job of the teacher is to design learning activities that develop procedural concepts and skills that are inherently tied to specific historical knowledge and contexts that develop a holistic awareness of students' understanding. Teachers can measure students' progress over a designated period (e.g. a semester, a year, or two years) by designing assessments that demonstrate mastery of historical skills and concepts rather than by simply testing students' ability to memorise historical knowledge.

Well-designed learning programs and activities allow progression to take place according to the individual student's abilities rather than at prescribed age or stage levels. Concepts are deepened and strengthened by continuous revisiting, and students work towards mastery of each skill or concept. A hierarchy of concepts and skills, linked to knowledge and understanding, can be derived from the Australian Curriculum to measure progression in history.

Indicators of learning progression in the Australian Curriculum: History

Indicators of learning progression, or improvement, are evident in the Australian Curriculum's yearly Achievement Standards. Action verbs describe students' cognitive development and articulate a hierarchical taxonomy of learning. For example, in Year 4, students recognise the significance of events in bringing about change, and in Year 5 they progress to being able to describe the significance of people and events/developments in bringing about change (See Table 1).

When measured over a designated period of time and reported as a continuum of development, (e.g. over two semesters, or over two years) it is possible to track students' progression in understanding the concepts of significance and continuity and change, as well as their ability to move from recognising to describing. Likewise, in Year 7, students 'describe the effects of change on societies, individual and groups', and in Year 8 they progress to 'explain[ing] the causes and effects of events and development' (See Table 2).

We should be mindful that progression levels can vary greatly within a class because individual students have different learning abilities and can therefore progress at different rates. A student's performance may not necessarily neatly align with year groups or chronological age.

For example, a Year 4 student who may be particularly interested and 'gifted' at learning history may be operating at Year 6 level in terms of their ability to analyse and evaluate historical sources and construct an argument. Alternatively, a student who has English as a second language (ESL) or foreign language (EFL) may understand the historical concepts but struggle with presenting their findings in written and/or spoken English. The progress of individual students should be accounted for by using differentiated identifiers that consider such variables.

Table 1 Learning progression from Year 4 to Year 5 (adapted from ACARA 2021)

By the end of Year 4 (age 10) a student will be able to:	By the end of Year 5 (age 11) a student will be able to:
Recognise the **significance** of events in bringing about change	**Describe** the **significance** of people and events/ developments in bringing about change
Explain how and why life **changed** in the past	**Identify** the causes and effects of **change** on particular communities
Identify aspects of the past that have **remained the same**	**Describe** aspects of the past that have **remained the same**
Describe the **experiences** of an individual or group in the past	**Describe** the **experiences** of different people in the past
Sequence information about events and the lives of individuals in **chronological order** with reference to key dates	**Sequence** information about events and the lives of individuals in **chronological order** using timelines
Develop questions about the past	**Develop** questions for a **historical inquiry**
Locate, **collect** and **sort** information from different **sources** to answer **questions**	
Identify and **select** a range of sources and **locate**, **compare** and use information to answer **inquiry** questions	**Identify** a range of **sources** and **locate**, **collect** and **organise** information related to an inquiry
Analyse sources to detect **points of view**	**Analyse** sources to determine their origin and purpose and to **identify** different **viewpoints**
Develop and **present** texts, including narrative recounts **using historical terms**	**Develop**, **organise** and **present** texts, particularly narrative recounts and descriptions, **using historical terms and concepts**

Note: red text, historical thinking concept; purple text, historical skill

Table 2 Learning progression from Year 7 to Year 8 (adapted from ACARA 2021)

By the end of Year 7 (age 13) a student will be able to:	By the end of Year 8 (age 14) a student will be able to:
Suggest reasons for **change and continuity over time**	**Recognise** and **explain** patterns of **change and continuity over time**
Describe the effects of **change** on societies, individuals and groups	**Explain** the **causes and effects** of events and developments
Describe events and developments from the **perspective** of different people who lived at the time	**Identify** the **motives and actions** of people at the time
Explain the role of groups and the **significance** of particular individuals in society	**Explain** the **significance** of individuals and groups and **how they were influenced by the beliefs and values of their society**
Identify past evens and developments that have been **interpreted** in different ways	**Describe** different **interpretations** of the past
Sequence events and developments within a **chronological framework**, **using** dating contentions to represent and measure time	**Sequence** events and developments within a **chronological framework** with **reference** to periods of time
Develop questions to **frame a historical inquiry** when researching	**Develop** questions to **frame a historical inquiry** when **researching**
Identify and **select** a range of sources and **locate**, **compare** and **use** information to **answer inquiry questions**	**Analyse**, **select** and **organise** information from **primary and secondary sources** and **use** it as **evidence** to **answer inquiry questions**
Examine sources to **explain points of view**	**Identify** and **explain different points of view in sources**
Identify their origin and purpose when **interpreting sources**	**Identify** their origin and purpose and **distinguish between** fact and opinion when **interpreting sources**
Use historical terms and concepts, **incorporate** relevant sources, and **acknowledge** their sources of information in developing texts and **organising and presenting their findings**	Use **historical terms and concepts**, and **evidence** identified in sources, and acknowledge their sources of information when **organising and presenting their findings**

Note: red text, historical thinking concept; purple text, historical skill

Measuring learning progression in history

Although curriculum documents provide teachers with guidelines and descriptors of learning progression at different stages in a continuum of learning, measuring a student's progression in history is not necessarily an easy task. Measurement of progression is usually achieved by using some form of assessment and then reporting the results. Unfortunately, there is no single formula or model for designing assessments to identify learning progression in history. The challenge is for educators to design appropriate assessment instruments that effectively reflect, measure and report on students' learning.

If the overall aim of learning is for students to improve, then assessment should be an instrument designed to identify, diagnose, and articulate improvement in relation to what the student could do before, rather than only reporting what the student can do at the time of assessment.

Researchers advise that diagnosis is best achieved through formative assessment that allows teachers to not only assess students' learning, but most importantly, assess the effectiveness of their teaching (see Carr & Counsell, 2014; William, 2011). The teacher should be able to identify students' strengths and weaknesses so that learners can improve and progress (Phillips, 2002). Assessment should not conflate attainment (or achievement) and progress; it should provide students and teachers with meaningful information on how they both can improve in the future.

Meaningful assessment descriptors should articulate the knowledge, conceptual understandings and skills that are typical for a learner to achieve at each level. These can be identified in assessments in the following ways:

1. **Benchmarks or standards** that are measurable criteria against which learning can be evaluated. These are often developed at a macro level by state or national assessment authorities. Benchmarks might be set for the level at which concepts or skills must be mastered in each grade. They might also be used to determine where a particular student, class, or school ranks in comparison to others. Meyer and Land (2006) call this point of mastery 'threshold concepts', which are determined to be central to a subject, and when understood by students, allow them to 'cross the threshold' of their understanding of that subject. Similarly, British history educator Alex Ford (2016) calls this a 'signpost'.
2. **Learning outcomes**, **objectives**, **goals** that are clear descriptions of what a learner is expected to be able to do, know about and/or value at the completion of learning. They describe the substance of learning and how its attainment will be demonstrated. These are often developed at the micro level, in schools, grades or classes for lessons or sequences of lessons. Threshold concepts or signposts are also valid measures of progression at the micro levels of class, grade and school.

The challenge is for teachers to develop progression models and assessment instruments that clearly define levels of proficiency that reflect an individual's level of achievement and the complexity of their learning.

References

Australian Curriculum, Assessment and Reporting Authority (ACARA) (2021). Retrieved 1 August, 2021 from http://australiancurriculum.edu.au/f-10-curriculum/humanities-and-social-sciences/

Carr, E., & Counsell, C. (2014). Using timelines in assessment. *Teaching History, 157*, 54–62.

Ford, A. (2016). *Progress and progression in history*. http://www.andallthat.co.uk/uploads/2/3/8/9/2389220/apf_-_progression_pack.pdf

Lee, P. J. (2005). Historical literacy: Theory and research. *International Journal of Historical Learning, Teaching and Research, 5*(1). https://www.researchgate.net/publication/242158846_Historical_Literacy_Theory_and_Research

Lee, P. J., & Ashby, R. (2000). Progression in historical understanding among students ages 7–14. In P. N. Stearns, P. Seixas, & S. Wineburg (Eds.), *Knowing, teaching, and learning history*, pp. 192–222. New York University Press.

Lee, P. J., Ashby, R., & Dickinson, A. K. (1996). Progression in children's ideas about history. In M. Hughes (Ed.), *Progression in learning*, pp. 51–81. Multilingual Matters.

Meyer, J. H. F., & Land, R. (2006). *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge*. Routledge.

Phillips, R. (2002). Historical significance: The forgotten key element. *Teaching History*, 106, 14–19.

William, D. (2011). *Embedded formative assessment*. Solution Tree Press.

The SWANs/ABLES Project: A set of resources developed collaboratively with teachers to support the teaching and learning of students with additional learning needs

Toshiko Kamei
University of Melbourne

https://doi.org/10.37517/978-1-74286-638-3_18

Toshiko Kamei is a research fellow at the Assessment Research Centre at the University of Melbourne. Her work has focused on projects to investigate the assessment and recognition of complex competencies in school-age students and adults. She has led projects that aim to develop resources to assess foundational skills in students with disability. Prior to her time at the Centre, she taught in primary, secondary and specialist settings in Canada, England, Japan and Australia. Toshiko's research interests include assessment of complex competencies, measurement theory, and raising the capacity of teachers to educate all students.

Abstract

The inclusion of students with additional learning needs in schooling is part of policy and practice in Australia. However, it has been well documented that teachers lack the resources and training to meaningfully include students with additional learning needs in the full range of learning in their classrooms. The SWANs (Students with Additional Needs) program of work aimed to fill this gap through developing assessments based on learning progressions to provide targeted information to support the teaching and learning of all students, including students with additional learning needs. The development and implementation of the SWANs/ABLES suite of resources illustrates how a practical assessment tool can support teachers to target the learning of students with additional learning needs regardless of their point of readiness to learn.

Introduction

Policies in Australian state education departments support the inclusion of students with additional learning needs in mainstream schools (e.g. Victorian Department of Education and Training) but the responsibility for implementing policy into practice lies with teachers (Sharma et al., 2012). However, teachers can lack the skills, resources and training to meaningfully include students with additional learning needs in classrooms (Senate Standing Committee on Education and Employment [Senate Committee], 2016; Victorian Equal Opportunity & Human Rights Commission [VEOHRC], 2012). The SWANs program of work aimed to fill this gap in knowledge through developing assessments based on learning progressions that provide teachers with targeted information to support the teaching and learning of all students including students with complex learning needs.

Research and development

Assessments based on learning progressions can provide teachers with information about what a student knows and what a student is ready to learn next to progress. An underlying assumption of learning progressions is that the skill to be measured can be structured in a way that describes increasing proficiency in skills and knowledge as students learn and develop. Thus, rather than describing a can-or-cannot conclusion about outcomes, assessments based on learning progressions aim to infer a student's level of achievement on a developmental continuum.

The SWANS program of work began in 2007 and was developed through two Australian Research Council (ARC) Linkage grants in partnership with the Victorian Department of Education and Training.

SWANs applied a method developed by Griffin (2007) that combined the work of Vygotsky (1980), Glaser (1981) and Rasch (1960) to build assessments based on learning progressions. Griffin (2007) equated Vygotsky's (1980) Zone of Proximal Development (a range of learning proficiency at which a student can progress with the support of a more capable other) to the point at which a student has a 0.50 chance of achieving a skill described in terms of criteria of success. In the Rasch model (Rasch, 1960), this point is estimated to be where the ability of the person is equal to the difficulty of the task. The work of Glaser (1981) on criterion-referenced interpretation of student performance was applied to describe students in terms of what they can do against criteria of performance quality.

SWANS aimed to expand the general curriculum to describe learning from a pre-intentional stage of learning to a stage where students are able to independently learn. In this way, it takes a strengths-based approach to describe students in terms of what they can do, regardless of their starting point for learning. Its intention is not to replace the general curriculum but to extend its access to all students. In this way, it describes emergent levels of learning in the following learning domains that were judged to be foundational skills for learning:

- **Communication**: the development of functional communication skills, building towards the use and understanding of social expectations about communication (Woods, 2010).
- **Literacy**: the development of the ability to make and interpret meaning using symbols (including pictures, signs, numbers, and text) leading towards early reading and writing (Woods, 2010).
- **Digital Literacy**: the development of the ability to interpret and use the language, symbols, and tools of digital technologies in a culturally appropriate manner. This includes learning to use technologies and using technologies to learn (White et al., 2017).
- **Numeracy**: the development of skills needed to notice, describe, understand and use numeracy information, including number and its operations, shape and pattern (Strickland et al. 2016).
- **Social Processes**: the development of skills to support social interaction, social responsibility, and a capacity to transcend social difficulties. These are the skills that help a student to learn both from and with others (Coles-Janess & Griffin, 2009).
- **Emotional Understanding**: the development of understanding about the experience and expression of emotions in self and others (Roberts, 2014).
- **Learning Skills**: the development of skills related to attention, memory, and executive functioning in school and classroom interactions, and that help students become more active and independent learners (Roberts, 2014).
- **Thinking Skills**: the development of strategies to actively participate in learning by using trial and error, evaluating outcomes, categorising, initiating activities and making choices. This leads to the development of skills involved in critical thinking such as predicting, planning, evaluating, and monitoring progress (Kamei, 2019).
- **Movement**: the development of the ability to achieve goals through strategic use of the movement capabilities of the body, enhancing agency, participation, and independence (Gale, 2018).

Initial framework development

The method of developing the assessment frameworks applied principles of validity based on the seminal work of Messick (1989) and employed procedures described by Wolfe and Smith (2007a; 2007b) and Wilson (2005).

The structure of the assessment frameworks was hierarchical. It defined the construct, or the skill to be measured. The construct is then broken down into strands that describe broad categories that are critical to the construct and then further into capabilities to describe key skills within each of the strands. Indicators were developed for each of the capabilities that are an indicative sample of a student's competence described as behaviours that students do, say, make or write. These indicators were then broken down into quality criteria that describe how well students can demonstrate competence in each indicator. This method of breaking the construct down into increasingly detailed levels of manifestation helped to ensure that the resulting assessment described the construct defined in the first step (Wilson, 2005).

A review of research was carried out to develop an initial theoretical assessment framework for each of the assessments consisting of a construct, capabilities, strands, indicators and quality criteria.

Collaboration with teachers

One important principle underlying the development of SWANs was that it was a tool for all teachers rather than specialists. Thus, an important aspect of development was the co-design process in collaboration with teachers. This phase of the methodology involved a series of workshops with subject matter experts (SMEs) who were teachers experienced in the teaching and learning of students with additional learning needs.

The SMEs reviewed the initial theoretical frameworks. They modified the initial framework and drafted additional indicators and quality criteria to reflect what the skills would look like in an educational setting in language that was accessible to all teachers. This process of review and drafting a pool of items ensured the assessments were interpretable and practical for all teachers and did not require specialist expertise. Moreover, it fostered a sense of shared ownership of the developed resources.

Large-scale field trial

In the next phase, large-scale field trials were carried out to collect student assessment data. The indicators were written as question stems and quality criteria as response options. This provided a set of questions in observation-based multiple-choice format for teachers to respond to based on their knowledge of their students accumulated through their regular interactions with them. Thus, students were not required to sit tests or to carry out specific tasks. Judgements of competence were made by teachers using their stored knowledge about students to choose responses based on evidence of what their students typically do, say, make or write.

The data were analysed using the Rasch partial credit model (Masters, 1982) to check the technical quality of the assessments. Estimates of difficulty were used to empirically order the quality criteria from lower to higher levels of difficulty. This information was used to derive a learning progression based on student assessment data.

SMEs were called on to review the empirically derived learning progression. They provided judgement as to where the key transitions were as students progressed in the domains of learning.

This step ensured that the resulting level statements described transitions that would be useful to teachers for planning their teaching and instruction. SMEs focused on recognising transitions that were observable key transitions that were useful for teachers while students progress from one level of learning to another. These were used as the basis to write level statements for the learning progressions.

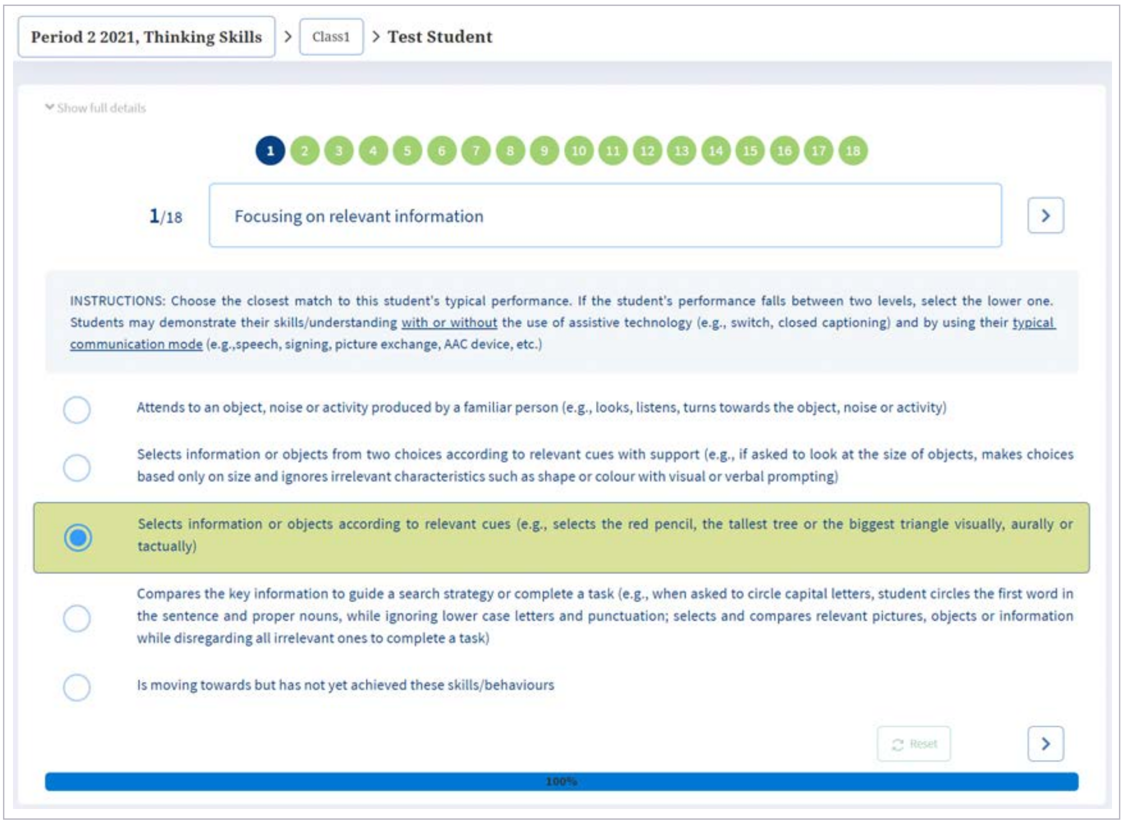
Instructional strategies

In the final step, SMEs were drawn on to review and write evidence-based instructional strategies mapped to each level of learning in the learning progressions to progress students from one level of learning to the next. Workshops took place where SMEs were presented with case studies of students and asked to make judgements on appropriate instructional strategies for them. Subsequently, the strategies were piloted with a separate group of SMEs who provided feedback on their practicability and applicability.

Structure of the assessment tools

The SWANs assessment instruments have subsequently been programmed to be delivered online in a questionnaire format. Figure 1 is an example of a SWANs item for Thinking.

Figure 1 SWANs item for Thinking



Teachers respond to the series of questions and a report is generated based on their responses. An example of a learning report is shown in Figure 2.

Figure 2 SWANs learning report nutshell statements

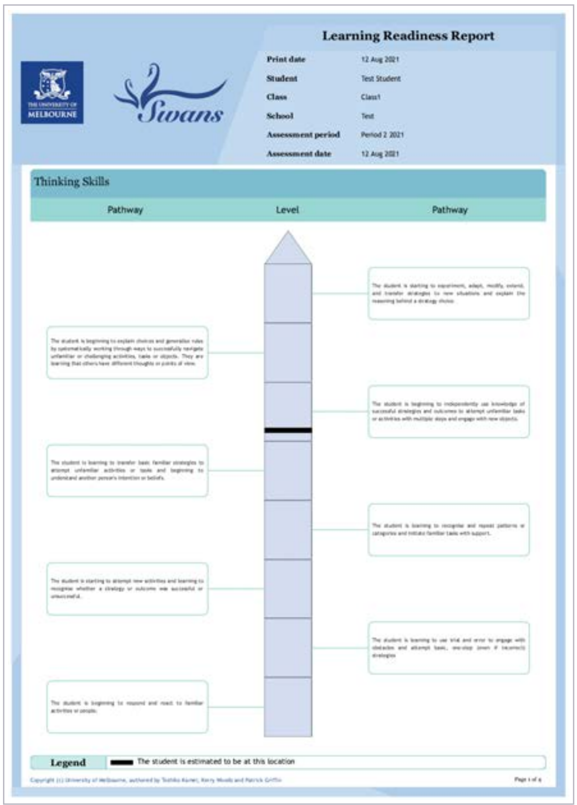
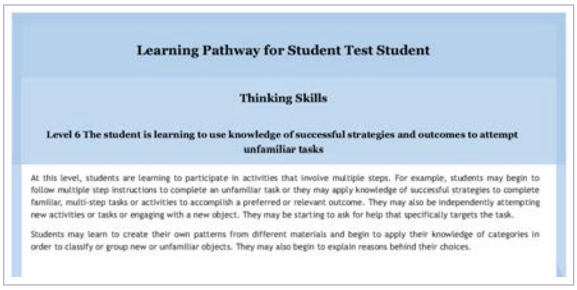


Figure 3 SWANS learning report extended statement



Learning Pathway for Student Test Student

Thinking Skills

Level 6 The student is learning to use knowledge of successful strategies and outcomes to attempt unfamiliar tasks

At this level, students are learning to participate in activities that involve multiple steps. For example, students may begin to follow multiple step instructions to complete an unfamiliar task or they may apply knowledge of successful strategies to complete familiar, multi-step tasks or activities to accomplish a preferred or relevant outcome. They may also be independently attempting new activities or tasks or engaging with a new object. They may be starting to ask for help that specifically targets the task.

Students may learn to create their own patterns from different materials and begin to apply their knowledge of categories in order to classify or group new or unfamiliar objects. They may also begin to explain reasons behind their choices.

Figure 4 SWANS teaching and learning strategies

Teaching and Learning Strategies for Student Test Student

Thinking Skills

Level 6 The student is learning to use knowledge of successful strategies and outcomes to attempt unfamiliar tasks

- Observe your student and maintain a record of activities the student finds enjoyable and motivating. Draw on information about the student's preferences, responsiveness, and support needs that has been shared by the student's parents/carers and/or by other teachers who know the student well. Draw on this information to motivate the student's participation at different times of the day.
- As appropriate for the student, draw on information and advice from support professionals (e.g., speech therapist, physiotherapist, occupational therapist) or disability-specific organisations to tailor activities for the student.
- As appropriate for the student, associate visual supports and verbal descriptions with objects and activities. These might include augmentative and alternative communication (AAC), key word signing, and picture symbols.
- Support your student's participation in multi-step activities by preparing clearly structured and sequenced activities that have a well-defined outcome. Use explicit instruction by, for example, demonstrating a step-by-step plan (strategy) for solving a problem specific to a task or activity. Implement explicit instruction in conjunction with other instructional components (e.g., the use of visual representations such as video modelling, or guiding your student to 'think aloud' about the steps in an activity).
- Practise monitoring and recording class goal completion (e.g., use a progress chart, sticker chart, calendar).
- Support the student to follow an established routine with a familiar adult and gradually decrease the level of support provided as your student learns to independently carry out multi-step tasks (e.g., setting up a classroom task or activity, finding and sorting the materials they need for a project of personal relevance and interest).
- Rearrange or modify instructional materials to improve your student's successful participation in new activities (i.e., provide scaffolding, examples, visual representations of steps in the activity, a daily/weekly schedule to guide the student through stages towards completion of a simple project). Remind the student of previous success in similar multi-step tasks.
- Ask questions that promote strategy use, self-instruction, self-monitoring, and self-evaluation (e.g., What comes next? Can you show me where you are on the visual schedule? 'How do you feel about _____ ?').
- Provide specific, task-oriented feedback to let the student know what s/he is doing well and what to revise.
- Before providing direct help, first ask your student to describe or explain the nature of the problem with a task and encourage him/her to imagine some possible solutions.
- Give the student a challenge such as an unfamiliar task and allow your student space, time, and a supported environment to think about ways to approach the task. Encourage the use of collaboration between a pair or small group of students to solve simple but challenging problems. Give the group examples of different ways to think about a problem, such as generating ideas and talking about preferences or opinions.

Figure 2 illustrates the learning progression for the learning domain by presenting 'nutshell statements' or brief descriptions of the competencies within each level of learning. The level within which a student is working is denoted by the dark black line.

Figure 3 shows an extended statement that describes in more detail the level within which a student is working. Teachers can use this information to set targeted short and long term goals for their student.

Figure 4 shows teaching and learning strategies that provide suggested teaching and learning strategies for teachers to use to progress their student from their present level of learning to the next.

SWANs and its connection to ABLES

The SWANS suite of materials were developed in conjunction with the ABLES (Abilities Based Learning and Education Support) work to strengthen its connection with the Victorian curriculum and enhance its use in Victorian schools. The ABLES version of the assessments links the foundational skills to the most relevant learning domains in the Victorian school curriculum (Victorian Curriculum and Assessment Authority [VCAA], 2021) and reports student progress in these terms.

Impact and further development

The SWANs and ABLES tools were subsequently programmed online to be disseminated nationally. Approximately 360 000 student assessments have been carried out to date by the SWANs/ABLES assessments.

Informing the curriculum

The SWANs resources informed the development of the Victorian Towards Foundation Curriculum (Underwood, 2020) that was developed in response to the Disability Standards for Education (Australian Government Department of Education and Training, 2005) that set out that all students should have access to curriculum on the same basis. Teachers of students with more severe disability commented that often, it was hard for them to 'see' their students in the curriculum (Underwood, 2020, p. 209). The SWANs research on interpersonal skills or social skills, communication, emotional skills, and cognitive or learning skills were particularly relevant to both the content and structure of the Towards Foundation Curriculum (Underwood, 2020).

Ongoing development

The SWANs work has been foundational in the development of further resources and work is ongoing. There have been two projects with the Victorian Department of Education and Training to adapt the SWANs suite of resources for use for students with additional learning needs in early childhood settings. This resulted in a set of assessments called Early ABLES. This has then further led to work to adapt the Early ABLES suite of resources for all children aged two to six years in funded kindergarten programs called the Early Years Assessment and Learning Tool. This work commenced in 2020 and is ongoing.

In addition, work has taken place with the Australian Curriculum and Assessment Authority (ACARA) to map the SWANs suite of resources to the national literacy and numeracy learning progressions. This was undertaken to ensure that the learning of all students, including students with additional learning needs, was included in the national learning progressions. This also led to further work to map the SWANs resources to the ACARA Critical and Creative Thinking and Personal and Social Capabilities Continua.

Conclusion

The SWANs program of work has continued since 2007 through two ARC Linkage grants in partnership with the Victorian Department of Education and Training. It drew together input from hundreds of assessment specialists, school leaders, teachers, specialist professionals, and curriculum leaders and is based on thousands of points of student assessment data. Through this process, all teachers now have the tools to understand students using the perspectives of teachers experienced in working with students with additional learning needs.

The SWANs materials were developed with the support of the Australian Research Council (ARC) as part of a Linkage partnership with the Centre for Advanced Assessment and Therapy Services and the University of Melbourne's foundation research partner, the Victorian Department of Education and Training. The Victorian Department of Education and Training is the development sponsor and owner of all rights in the ABLES Tools. The ABLES Tools were derived from the Students with Additional Needs (SWANs) assessment and reporting materials which are owned by the University of Melbourne.

References

Australian Government Department of Education and Training (2005). *Disability standards for education*. https://docs.education.gov.au/system/files/doc/other/dse-fact-sheet-2-dse_0.pdf

Coles-Janess, B., & Griffin, P. (2009). Mapping transitions in interpersonal learning for students with additional needs. *Australasian Journal of Special Education, 33*(2), 141–150. https://doi.org/10.1375/ajse.33.2.141

Gale, L. (2018). *The acquisition of goal-directed movement: Developing and reporting a measure of learning readiness for use in classroom settings with students with motor performance needs* [Doctoral dissertation, University of Melbourne]. Minerva Access. http://hdl.handle.net/11343/217267

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36*(9), 923–36. http://dx.doi.org/10.1037/0003-066X.36.9.923

Griffin, P. (2007). The comfort of competence and the uncertainty of assessment. *Studies in Educational Evaluation, 33*(1), 87–99. https://doi.org/10.1016/j.stueduc.2007.01.007

Kamei, T. (2019). *Thinking to learn: An investigation of assessing and teaching thinking skills for students with disability* [Doctoral dissertation, University of Melbourne]. Minerva Access. http://hdl.handle.net/11343/224151

Masters, G. (1982). A rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174. https://doi.org/doi:10.1007/BF02296272

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). MacMillan Publishing.

Rasch, G. (1960) *Probabilistic models for some intelligence and attainment tests*. Neilson & Lydiche.

Roberts, E. (2014). *Design and validation of an instrument to assess development of intrapersonal learning skills for students with an intellectual disability and students with autism spectrum* [Doctoral dissertation, University of Melbourne]. Minerva Access [Abstract only]. https://minerva-access.unimelb.edu.au/handle/11343/39767

Senate Standing Committee on Education and Employment (2016). *Access to real learning: The impact of policy, funding and culture on students with disability*. Senate Printing Unit, Parliament House.

Sharma, U., Loreman, T. & Forlin, C. (2012). Measuring teacher efficacy to implement inclusive practices. *Journal of Research in Special Educational Needs, 12*(1), 12–21. https://doi.org/doi: 10.1111/j.1471-3802.2011.01200.x

Strickland, J., Woods, K., & Pavlovic, M. (2016, November 27–December 1). *Assessing and understanding early numeracy for students with additional learning needs* [Conference paper]. Australian Association for Research in Education, Melbourne. https://www.researchgate.net/publication/318897270_ASSESSING_AND_UNDERSTANDING_EARLY_NUMERACY_FOR_STUDENTS_WITH_ADDITIONAL_LEARNING_NEEDS

Underwood, K. (2020). A curriculum for students with additional needs. In Griffin P. & Woods, K. (Eds.) *Understanding students with additional needs as learners*. (pp. 209–216).Springer International Publishing. http://doi.org/10.1007/978-3-030-56596-1_14

Victorian Curriculum and Assessment Authority. (2021). *The Victorian curriculum*. http://victoriancurriculum.vcaa.vic.edu.au/

Victorian Department of Education and Training. (2021). *School operations: Students with disability*. https://www2.education.vic.gov.au/pal/students-disability/policy

Victorian Equal Opportunity & Human Rights Commission. (2012). *Held back: The experiences of students with disabilities in Victorian Schools*. http://www.humanrightscommission.vic.gov.au/media/k2/attachments/1404-VEOHRC_HeldBack_-_StudentwithDisabilityReportW3.pdf

Vygotsky, L. S. (1980). *Mind in society: The development of higher mental processes*. Harvard University Press.

White, E. H., Woods, K., & Poed, S. (2017, July 2–7). *The assessment and development of digital literacy in students with vision impairment and additional learning needs: Preliminary findings from a current PhD study* [Conference paper]. 9th ICEVI European Conference, Bruges. http://www.icevi-europe.org/files/2017/empowered-by-dialogue/icevi-europe-2017-proceedings.pdf

Woods, K. (2010). The design and validation of measures of communication and literacy to support the instruction of students with learning disabilities. [Doctoral dissertation, University of Melbourne]. Minerva Access [Abstract only]. http://hdl.handle.net/11343/35468

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Lawrence Erlbaum Associates.

Wolfe, E.W., & Smith, E.V., Jr. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part I – Instrument development tools. In E.V. Smith, Jr. & R. M. Smith (Eds.). *Rasch measurement: Advanced and specialized applications* (pp. 202 – 242). JAM Press.

Wolfe, E.W., & Smith, E.V., Jr. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part II – Validation activities. In E.V. Smith, Jr. & R. M. Smith (Eds.). *Rasch measurement: Advanced and specialized applications* (pp. 243 – 290). JAM Press.

PROGRAM

Day 1: 16 August

Keynote

How education gets in the way of learning

10:00am – 11:00am

Professor Geoff Masters AO, ACER

General principles of learning progressions

11:15am – 12:15pm

Charlotte Waters, ACER

Developing an assessment of oral language and literacy: Measuring growth in the early years

12:30pm – 1:30pm

Dr Dan Cloney and Kellie Picker, ACER

30-minute break

This time without 'feeling': Children's intuitive theories of art as a logical basis for learning progression in visual arts

2:00pm – 3:00pm

Dr Karen Maras, University of New South Wales

Keynote

Karmel Oration
Excellent progress for all: A function of year-level curriculum or evidence-based learning progressions?

3:15pm – 4:15pm

Professor Dianne Siemon, RMIT

Day 2: 17 August

Learning progressions as models and tools for supporting classroom assessment

10:00am – 11:00am
Associate Professor Alicia C Alonzo, Michigan State University

Applying empirical learning progressions for a holistic approach to evidence-based education: SWANs/ABLES

11:15am – 12:15pm
Dr Emily White, University of Melbourne

Evidencing creativity and curiosity in IB schools

12:30pm – 1:30pm
Dr Sarah Richardson and Dr Sladana Krstic, ACER

30-minute break

Exploring excellence in Indigenous education in Queensland secondary schools

2:00pm – 3:00pm
Dr Marnee Shay, Dr Jodie Miller and Dr Suraiya Hameed, University of Queensland

How might we identify and measure learning progression in History?

3:15pm – 4:15pm
Dr Louise Zarmati, University of Tasmania

Day 3: 18 August

Keynote

Accountable assessment

10:00am – 11:00am
Professor Richard Lehrer, Vanderbilt University

Using assessment data to improve equity: How teachers use insights from the Scottish National Standardised Assessments

11:15am – 12:15pm
Dr Sarah Richardson and Dr Sladana Krstic, ACER

Interpreting learning progress using assessment scores: What is there to gain?

12:30pm – 1:30pm
Dr Nathan Zoanetti, ACER

30-minute break

Leading system transformation: A work in progress

2:00pm – 3:00pm
Gregory B Whitby AM, Maura Manning and Dr Gavin Hays, Catholic Education Diocese of Parramatta

Making excellent progress in early reading: How can the identification of essential skills and concepts help?

3:15pm – 4:15pm
Dr Dara Ramalingam, Prue Anderson, Sandra Knowles, Danielle Anzai and Greta Rollo, ACER

Day 4: 19 August

Keynote

Rethinking measurement for accountable assessment

10:00am – 11:00am
Professor Mark Wilson, University of California, Berkeley, and University of Melbourne

Supporting science teaching practice with learning progressions

11:15am – 12:15pm
Professor Erin Furtak, University of Colorado Boulder

Identifying and monitoring progress in collaboration skills

12:30pm – 1:30pm
Dr Claire Scoular, ACER

30-minute break

The SWANs/ABLES Project: A resource developed in, by and for teachers to support the teaching and learning of students with additional learning needs

2:00pm – 3:00pm
Dr Toshiko Kamei

Reporting student progress: What might it look like?

3:15pm – 4:15pm
Hilary Hollingsworth and Jonathan Heard, ACER; Anthony Hockey, St Mary's Primary School and Tegan Knuckey, Balwyn High School, Victoria

Masterclass session 1: 20 August

Keynote

Learning progressions in reading and mathematics

10:00am – 2:00pm

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

Masterclass session 2: 25 August

Keynote

Learning progressions in reading and mathematics

10:00am – 2:00pm

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