



ARC Linkage project (LP180100046)

Improving teacher assessment capability
using scaled annotated exemplars of
achievement standards
in online moderation

Project Research Team

Associate Professor Lenore Adie	Australian Catholic University
Professor Claire Wyatt-Smith	Australian Catholic University
Associate Professor Stephen Humphry	The University of Western Australia
Professor Michele Haynes	Australian Catholic University
Professor David Andrich	The University of Western Australia
Professor Christopher DeLuca	Queen's University, Ontario, Canada
Mr Terry Gallagher	Queensland Department of Education
Project Manager Jo Macri	Australian Catholic University

Advisor: Professor You-Gan Wang Australian Catholic University

Technology: Mr Alexander Chen Australian Catholic University

Partner Organisations

Queensland Department of Education
WA School Curriculum and Standards Authority
Catholic Education Western Australia Limited

This research was funded by the Australian Research Council (LP180100046)

Acknowledgement

Our sincere gratitude goes to the teachers who worked alongside us in this research project during one of the most difficult times for schooling in the 21st century - the impact and consequences of the COVID pandemic. We greatly appreciate your professionalism and dedication to improving assessment processes and practices for teachers and students.

This project was funded by the Australian Research Council Linkage Grant Scheme and has been supported by Partner Organisations the Queensland Department of Education, Western Australian School Curriculum and Standards Authority, and Catholic Education Western Australia Limited.



Background



psychometrically calibrated work samples
(Heldsinger & Humphry, 2010, 2013)
+
cognitive commentaries

marking guides

Australian Curriculum
Achievement Standards
+
State descriptors of A-E standards,
annotated work samples



exemplars
+
verbal
descriptions

Dependability of
judgement



Sadler (1987)
Wyatt-Smith, (1995)



Project aim

Systematically investigate the use of exemplars to make comparable judgements against achievement standards, regardless of school location, utilising approaches that are informed by the metacognition of judgement-making.



- mixed-method research design
- Queensland and Western Australia
- English, science, maths, and in WA only, religious education
- middle years of schooling (Yrs 4, 6 and 8)



Four project objectives

To improve teacher consistency of A-E judgements of student achievement through development of statistically scaled exemplars of achievement standards with explanations of teacher judgements.

To use statistically scaled exemplars of achievement standards and expert judgement to identify A-E standards of performance.



To develop an application of digital technology that enables teachers to grade and moderate students' work using the annotated scaled exemplars regardless of geographic location.





















To collaborate with Industry Partner Organisations to develop teachers' evaluative knowledge and expertise when assessing student achievement in the middle years of schooling.



Limitations

Impact of COVID pandemic on teaching workforce and workload; school closures

Participant numbers and spread across regions

Timeline of project events

Timeline		Yr1 (2/9/19 – 1/9/20)		Yr2 (2/9/20 – 1/9/21)		Yr3 (2/9/21 – 28/04/23)				
		Jan- June 2019	July- Dec 2019	Jan- June 2020	July- Dec 2020	Jan- June 2021	July- Dec 2021	Jan- June 2022	July- Dec 2022	Jan- Apr 2023
	<ul style="list-style-type: none"> Original project duration MIA Ethics Approval (ACU) 									
Stage 1	<ul style="list-style-type: none"> Preparation for Stage 1 inc. identification/development of assessment tasks Recruitment Term 1, 2020 – Term 3, 2021 Sample collection Term 3, 2020 – Term 3, 2021 	 								
Stage 2	<ul style="list-style-type: none"> Preparation for Stage 2 Pairwise comparison Term 3-4, 2021 									
Stage 3	<ul style="list-style-type: none"> Preparation for Stage 3 Standard-setting procedure Term 4, 2021 									
Stage 4	<ul style="list-style-type: none"> Preparation for Stage 4 Development of cognitive commentaries Term 1, 2022 Finalisation of cognitive commentaries Term 2, 2022 									
Stage 5	<ul style="list-style-type: none"> Preparation for Stage 5 Trial of online moderation platform Term 2-3, 2022 Expert judges validation of scoring Term 3, 2022 									
	<ul style="list-style-type: none"> Data analysis, publications, dissemination Term 1, 2022 – Apr, 2023 									

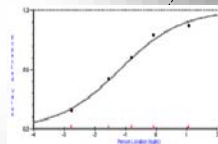
Key: Original timeline 
Deferred timeline 

Project stages

Stage 1: 2020/2021

Collection of assessment tasks
(Yrs 4, 6, 8 English narrative,
maths & science investigations)

- 1155 assessments
- 85 teachers



Stage 3: 2021

Standard-setting

- 89 teachers and 12 PPOs
- 535 judgements



Stage 2: 2021

Online pairwise comparison

- 755 assessments
- 94 teachers and 6 PPOs



Stage 4: 2022

Development of cognitive commentaries

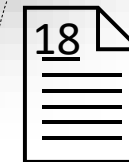
- 46 teachers and 11 PPOs
- 162 commentaries
- 14 online meetings



Stage 5: 2022

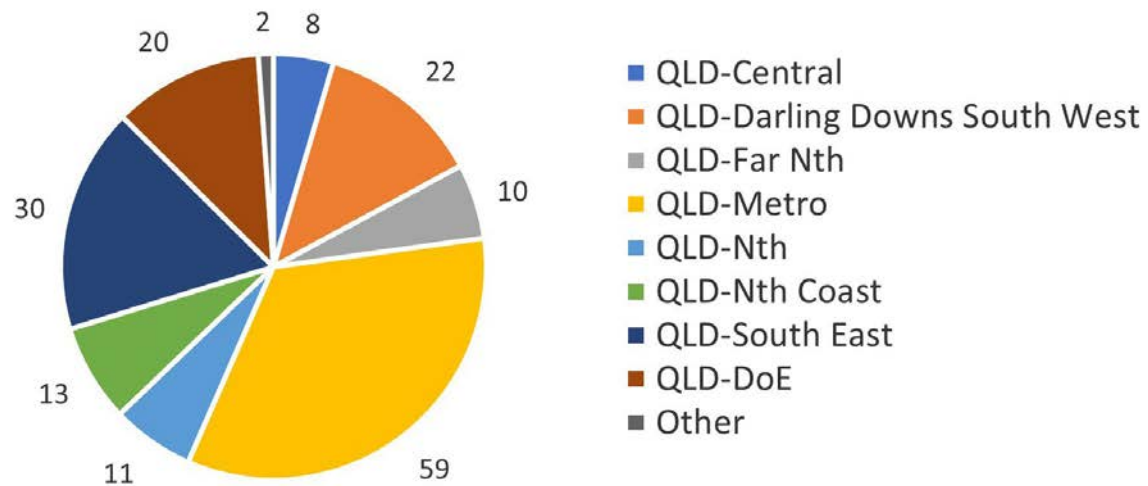
Trial of scaled
exemplars with
associated
cognitive
commentaries

- 56 teachers
and 18 PPOs
- 43 scaled
exemplars

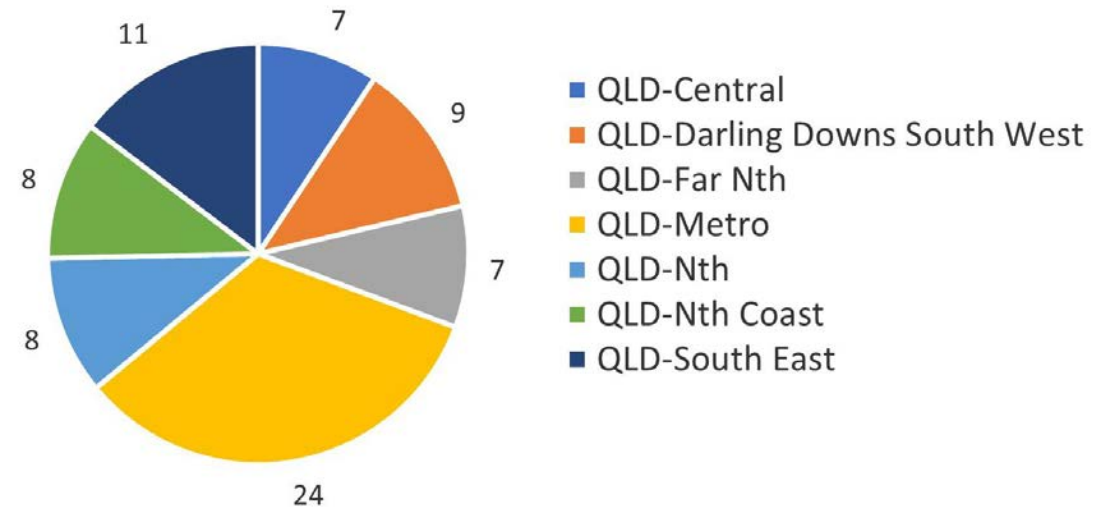


Participation by Queensland schools across the project

Participants per regional area/Dept



Schools by regional area



Number of project participants across regions and stages

	Stage 1		Stage 2			Stage 3			Stage 4a			Stage 4b		Stage 5		
	Schools	Teachers	Schools	Teachers	PPOs	Schools	Teachers	PPOs	Schools	Teachers	PPOs	Schools	Teachers	Schools	Teachers	PPOs
Central Queensland	1	2	5	6		5	6		4	5		4	4	4	5	
Darling Downs South West	6	11	5	11		5	11		3	6		3	5	4	9	
Far North Queensland	4	5	3	4		3	4		2	2		2	2	3	5	
North Queensland	5	5	6	9		5	8		2	2		2	2	1	1	
Metropolitan	12	31	10	32		11	32		8	18		7	17	11	21	
North Coast	5	8	7	11		6	9		1	1		0	0	1	1	
South East	6	23	7	21		6	19		3	12		2	10	6	14	
TOTAL	39	85	43	94	6	41	89	12	23	46	11	20	40	30	56	20

Number of participants within stages

	Stage 1	Stage 2	Stage 3	Stage 4	Stage 5
Stage 1	19	4	27	13	21
Stage 2		5	18	4	1
Stage 3			4	1	
Stage 4				1	4
Stage 5					30

Note: one participant participated in stages 1, 2 and 4.

- Participation in 1 stage
- Participation in 2 stages
- Participation in 3 stages
- Participation in 4 stages
- Participation in 5 stages

Reasons for participant withdrawal

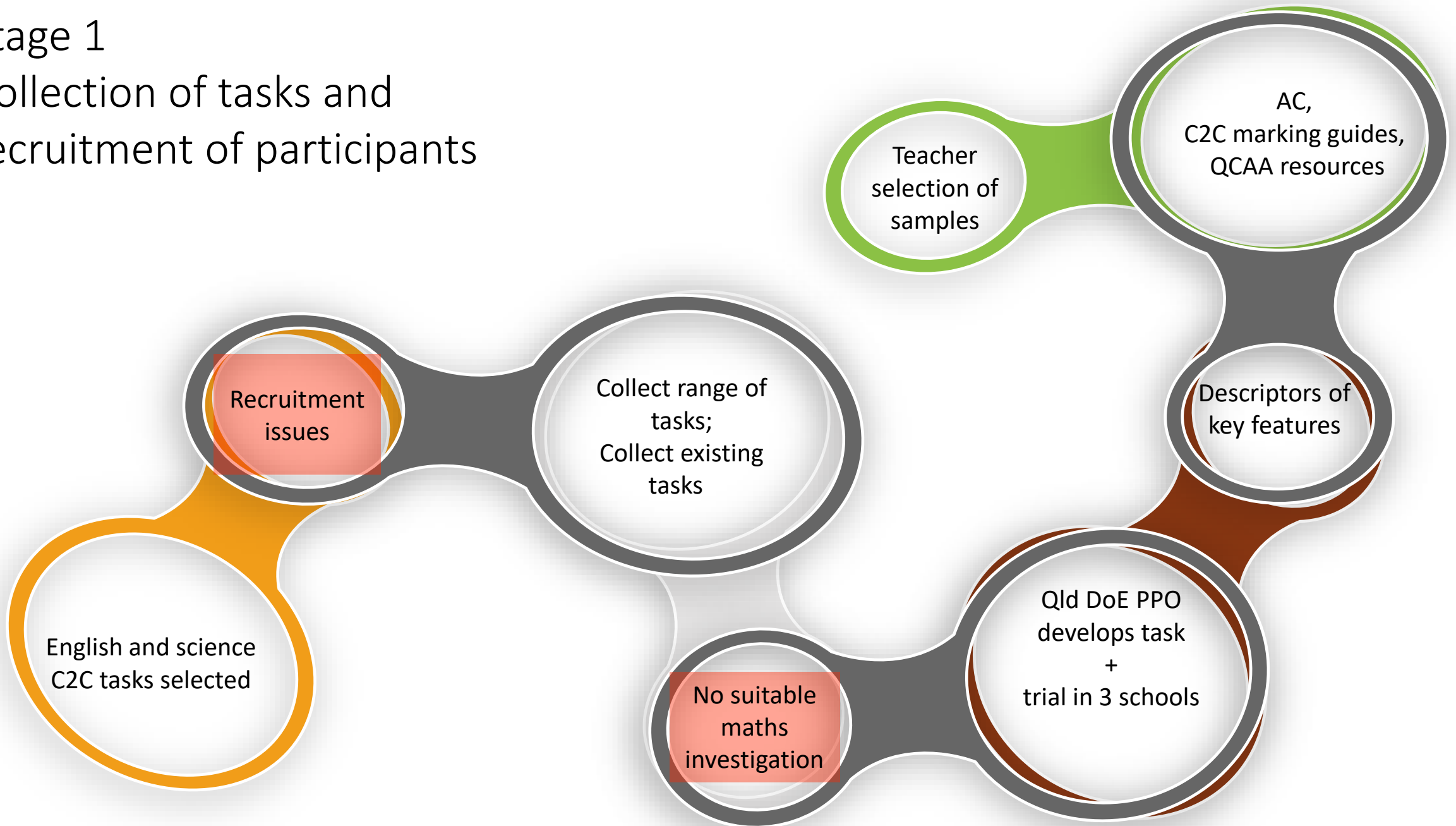
Withdrawal reason	Total
Workload	23
Covid-19	13
Flood	2
Change in role	16
On leave	8
Inability to participate	8
No response/reason provided	26
	96

Reasons for school non-participation

Non-participation reason	Total
Workload	9
Covid-19	4
Inability to participate	2
No reason provided	5
	20

Stage 1

Collection of tasks and recruitment of participants



Criteria for sample selection for pairwise comparison

A cross-section
of schools and
regions

Equal
distribution of
A-E grades

Legibility of
the sample

Stage 1: Number of Queensland teachers and samples per discipline

Region	No. of participating teachers			
	English	Maths	Science	Total
Central Queensland	0	2	0	2
Darling Downs South West	4	5	3	12*
Far North Queensland	3	1	1	5
North Queensland	4	1	0	5
Metropolitan	4	16	13	33*
North Coast	4	1	3	8
South East	5	5	13	23
Total	24	31	33	88

Discipline	Year Level	Samples submitted	Total
English	Yr4	188	465
	Yr6	162	
	Yr8	115	
Maths	Yr4	165	335
	Yr6	69	
	Yr8	101	
Science	Yr4	128	355
	Yr6	139	
	Yr8	88	
Total			1155

Note. * 3 teachers participated in more than 1 discipline. 85 teachers in total.

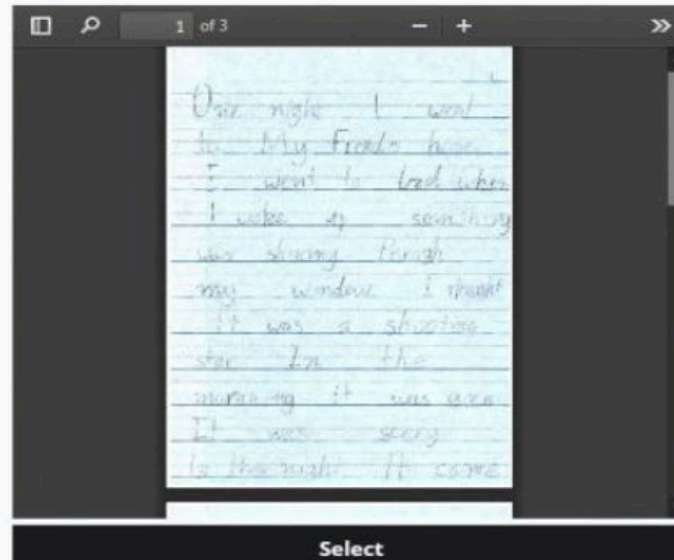
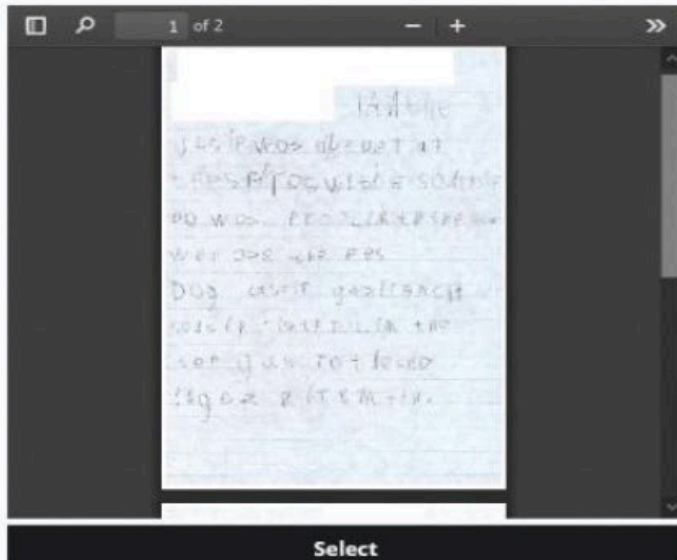
Demo ARC Narrative Yr 4 Yr 6

Team ARC Project

Description ARC Linkage Project Narrative Year 4 and 6 comparisons Demo.

Judging in Progress 40 pairs left to judge

Go Fullscreen



Which is better? Authorial

Next: Which is better? Conventions

Click on the sample which you believe is better

Stage 2

Pairwise comparison

Stage 2 Pairwise comparison: Selected samples

WA de-identified English and science scaled samples from across Years 3-7 and scored by WA teachers were added to the Queensland pool. These samples provided the means to establish comparability across states.

Breakdown of teacher A-E grades of selected Queensland samples

Grade	English			Maths			Science			Total
	Yr4	Yr6	Yr8	Yr4	Yr6	Yr8	Yr4	Yr6	Yr8	
A	16	15	9	10	10	6	12	11	14	103
B	16	16	14	10	9	6	14	11	17	113
C	19	13	14	10	16	8	14	11	11	116
D	8	8	10	11	11	7	8	11	8	82
E	1	2	3	9	4	1	2	6	0	28
Total	60	54	50	50	50	28	50	50	50	442

Number of Queensland and WA samples used in pairwise comparison

Pairwise Project	Qld			WA	Total
	Yr4	Yr6	Yr8	Yr unknown	
English Yr4/6	60	54	0	115	229
English Yr8	0	0	50	50	100
Maths Yr4/6	50	50	0	0	100
Maths Yr8*	0	10	28	0	38
Science Yr4/6	50	50	0	88	188
Science Yr8**	0	50	50	0	100
Total	160	214	128	253	755

*10 Yr 6 maths A-C samples were added to increase Yr 8 maths

**50 Yr 6 science A-C samples were added to increase Yr 8 science

Stage 2 Pairwise comparison: Methodology

Year 4 and Year 6 samples were combined in the pairwise process to form a common scale to enable comparison across year levels.

Scaled English and science exemplars (Years 3-7) scored by WA teachers were included in the pairwise comparison to establish comparability across states.

120 pairs were available for comparison within each discipline (English, maths, and science) and within grouped year levels.

Pairs generated randomly from the list of all pairs of performances.

A design was constructed in which each performance was compared with other performances twice in almost all cases.

One sample remained stable over four pairs to reduce the cognitive load for judges.

After four pairs, two new samples were presented.

Judges worked individually.

Holistic judgements were made on which of 2 performances displayed more advanced knowledge and skills.

Participants completed 40-120 comparisons.

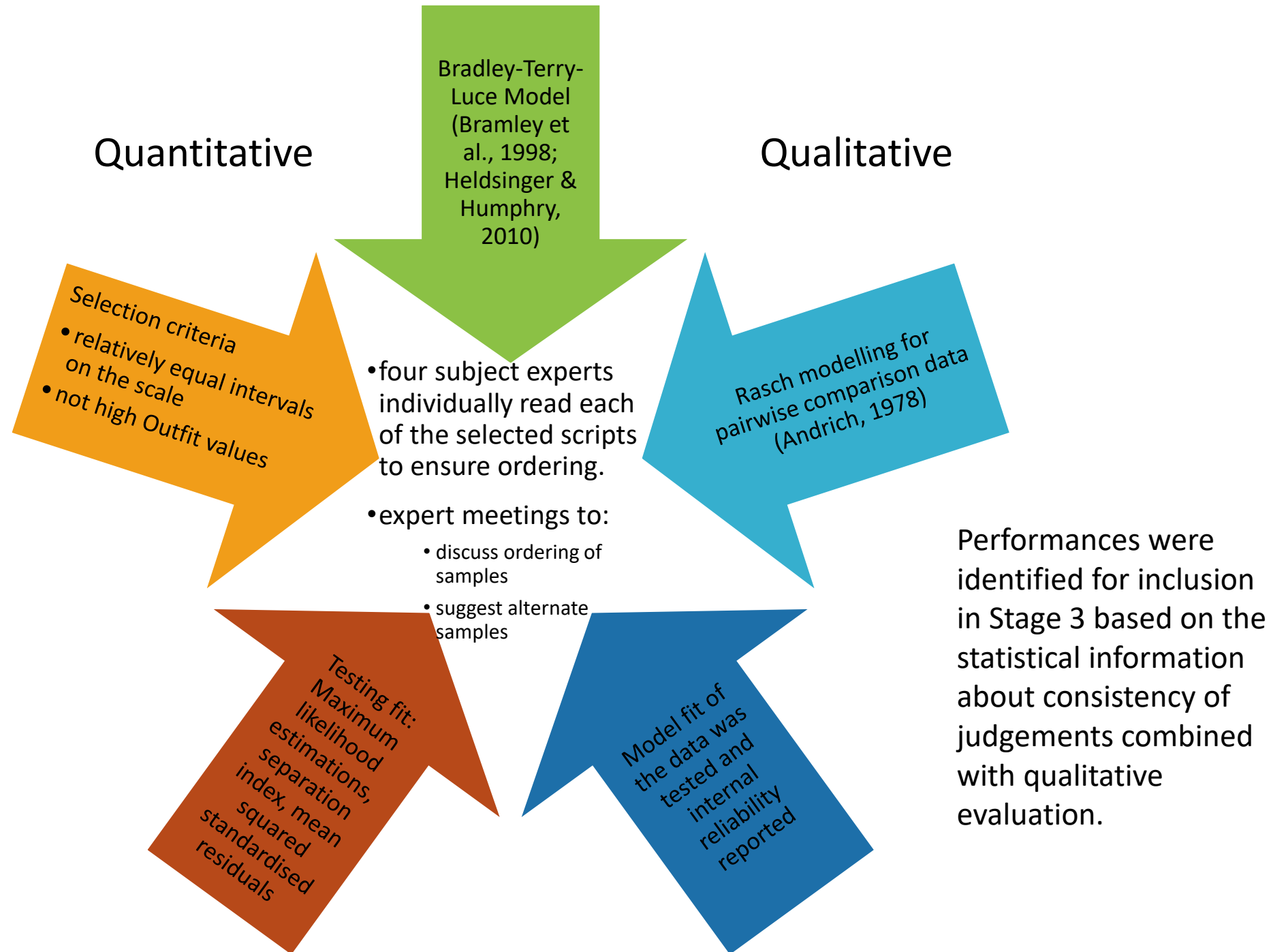
Number of pairwise comparison judgements made per year level grouping and discipline.

Pairwise project	Total
Yr4/6 English	4,022
Yr8 English	1,996
Yr4/6 Science	3,740
Yr8 Science	1,448
Yr4/6 Maths	4,660
Yr8 Maths	1,132
Total	16,998

Number of Queensland participants and average percent completion per year level and discipline.

Yr Level	Discipline	Participants	Average completion %
Yr4	English	12	74
	Maths	13	61
	Science	12	74
Yr6	English	9	86
	Maths	14	83
	Science	12	65
Yr8	English	10	83
	Maths	8	59
	Science	10	60
Total		100	

Stage 2
Pairwise
comparison:
Data
analysis



Stage 2 Pairwise comparison: Findings

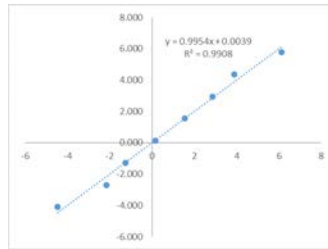
The linear line of best fit for final set of performances

A subset of 7-8 performances was selected for each year level and discipline.

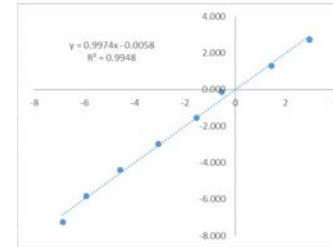
These constituted the referent work samples for Stage 3.

Only samples submitted for the corresponding year level were selected.

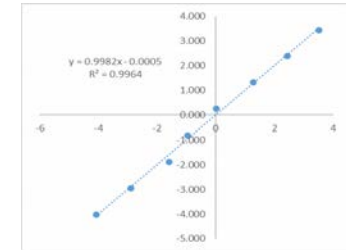
English Year 4



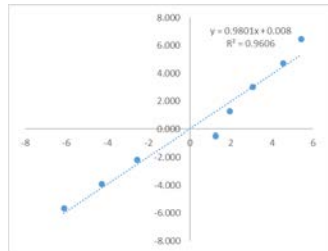
Maths Year 4



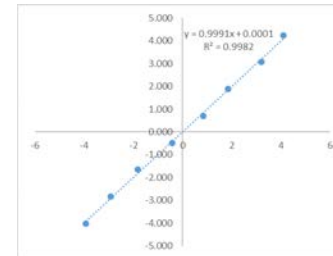
Science Year 4



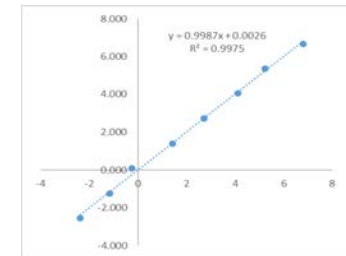
English Year 6



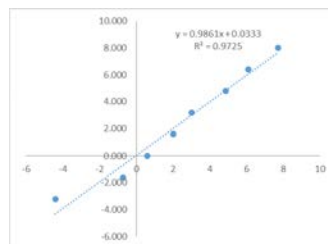
Maths Year 6



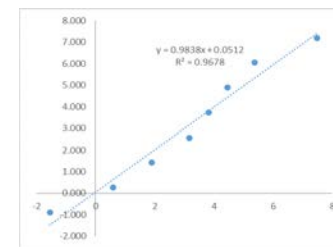
Science Year 6



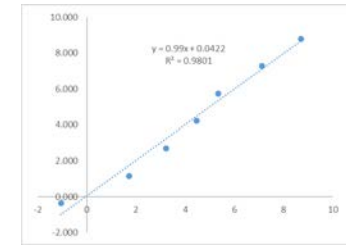
English Year 8



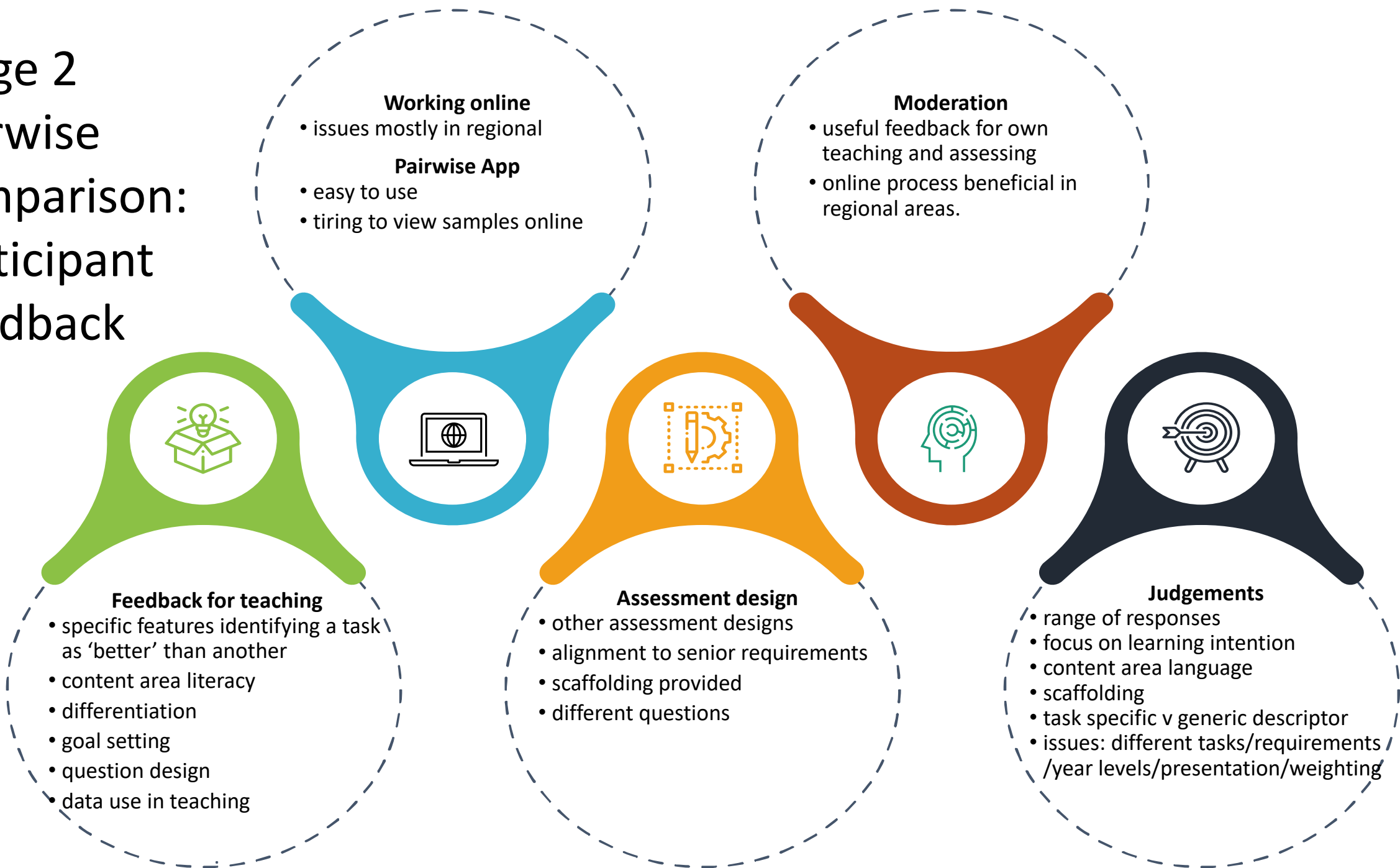
Maths Year 8



Science Year 8



Stage 2 Pairwise comparison: Participant Feedback



The score will be displayed at the top left below the A descriptor tab, and the B standard descriptor will be automatically displayed for you on the right side of your screen

Stage 3
Standard-
setting

ARC LINKAGE - RESEARCH PROJECT
brightpath Dashboard Projects Reports Manage Jo Macri Logout Knowledge Base Report An Issue

Projects / Assess: Year 4 Science Back to Project List

A Level Year 4 Science 320
B Level Year 4 Science
C Level Year 4 Science
D Level Year 4 Science
E Level Year 4 Science

Marked 1 of 5 students

B Level Year 4 Science Notes

B level

Work at this level typically shows accurate and appropriate measurements, observations and representation of results/data. Students will typically identify variables to be changed and controlled to make the investigation a fair test. Students will mostly use scientific understanding, language and representations to justify predictions, explain/justify inferences and results, and draw appropriate conclusions.

Professional judgement

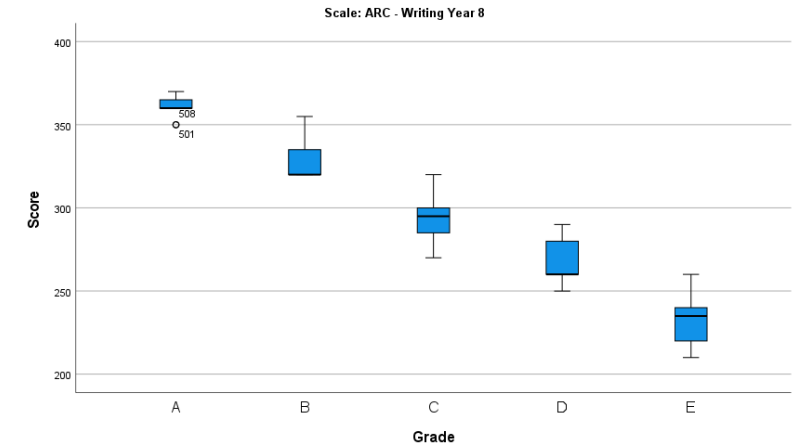
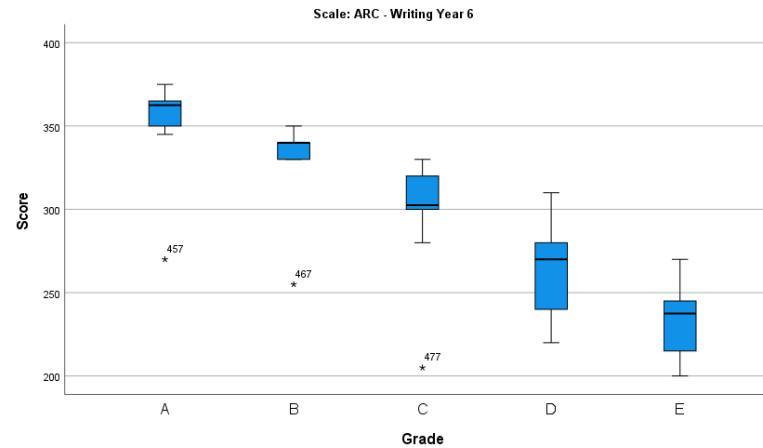
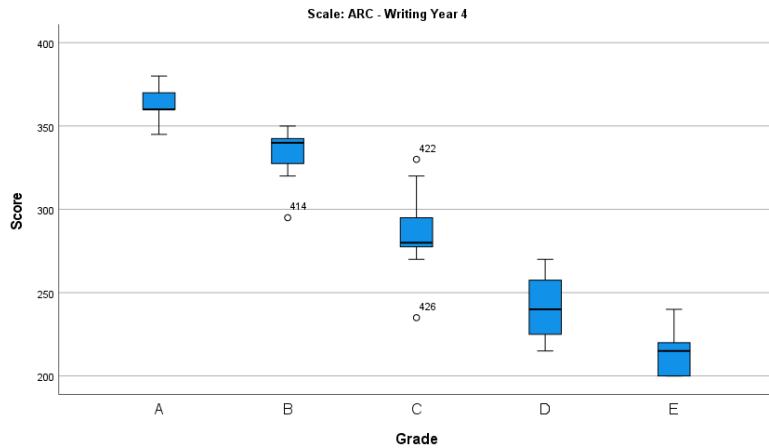
In making your judgement, draw on your professional judgement and experience to consider the degree to which performances show higher-order understanding and scientific investigation capabilities.

Terms and Conditions of Use © 2021 Pairwise Pty Ltd.

Stage 3 Standard Setting: Data analysis

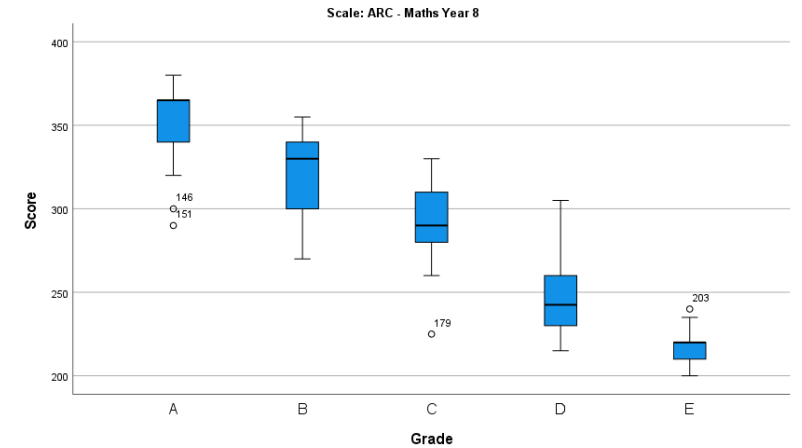
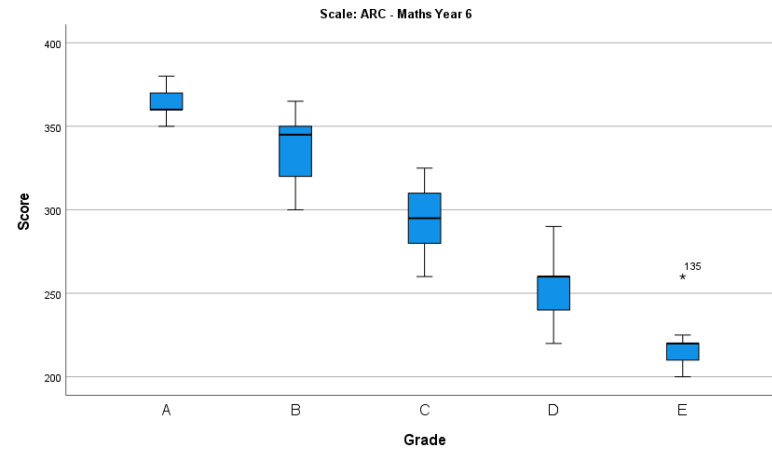
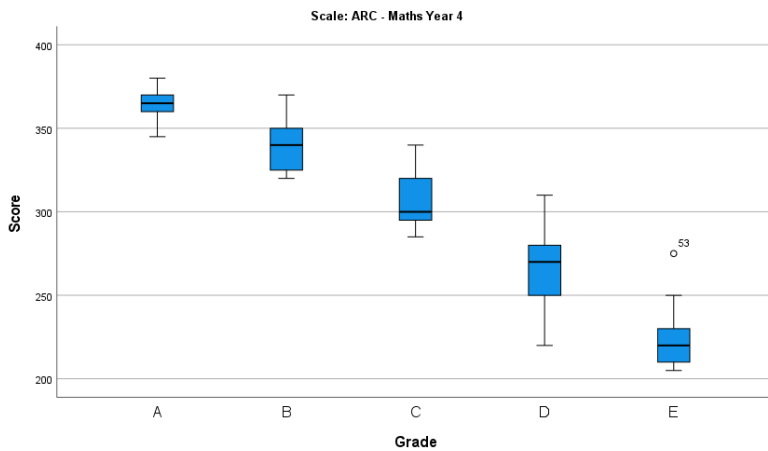
- Average, median and standard deviation of the scale locations for each standard across judges
- Face validity and utility of exemplars
 - the relation of average locations was compared across year levels (4, 6 and 8) by constructing single scales that span the year levels
- In most cases the mean and median scores were the same

Stage 3 Standard Setting: Findings English



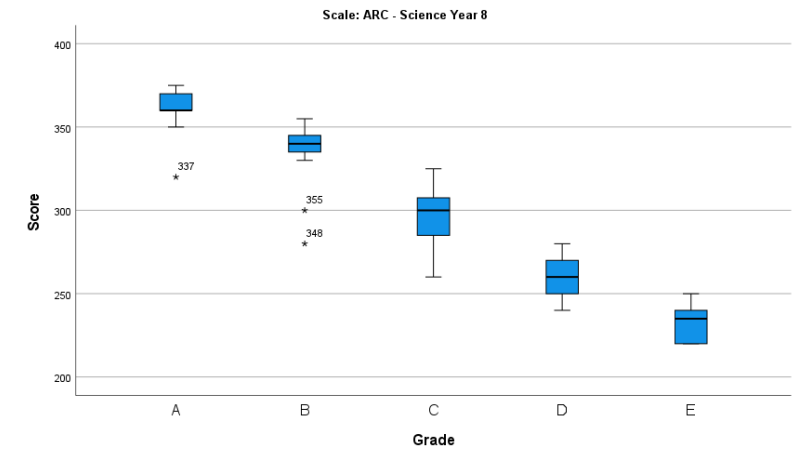
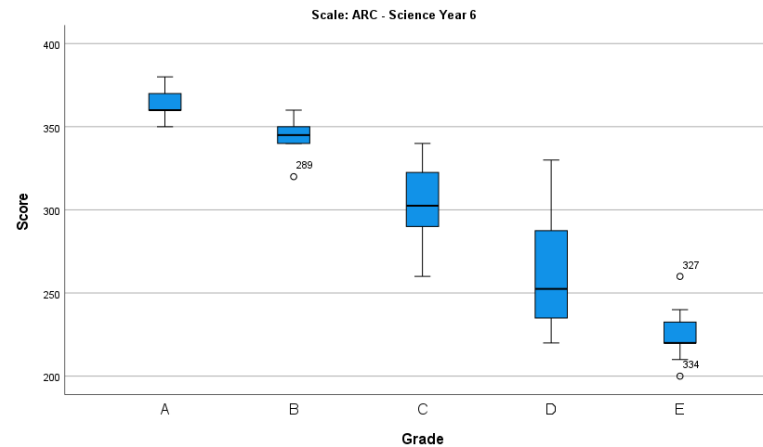
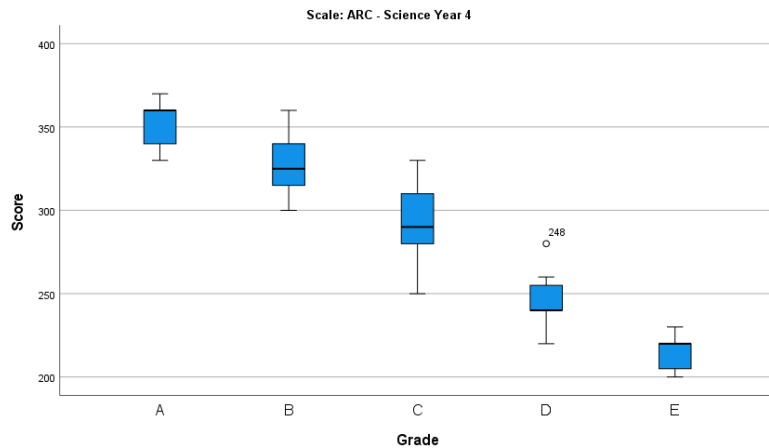
1. Median scale location for A level is only somewhat higher in Year 6 compared to Year 4
2. Relatively large range of scores selected for Year 4 D level
3. Significant variation for Year 6 D and E levels
4. 3 outliers in Year 6 are quite extreme

Stage 3 Standard Setting: Findings maths



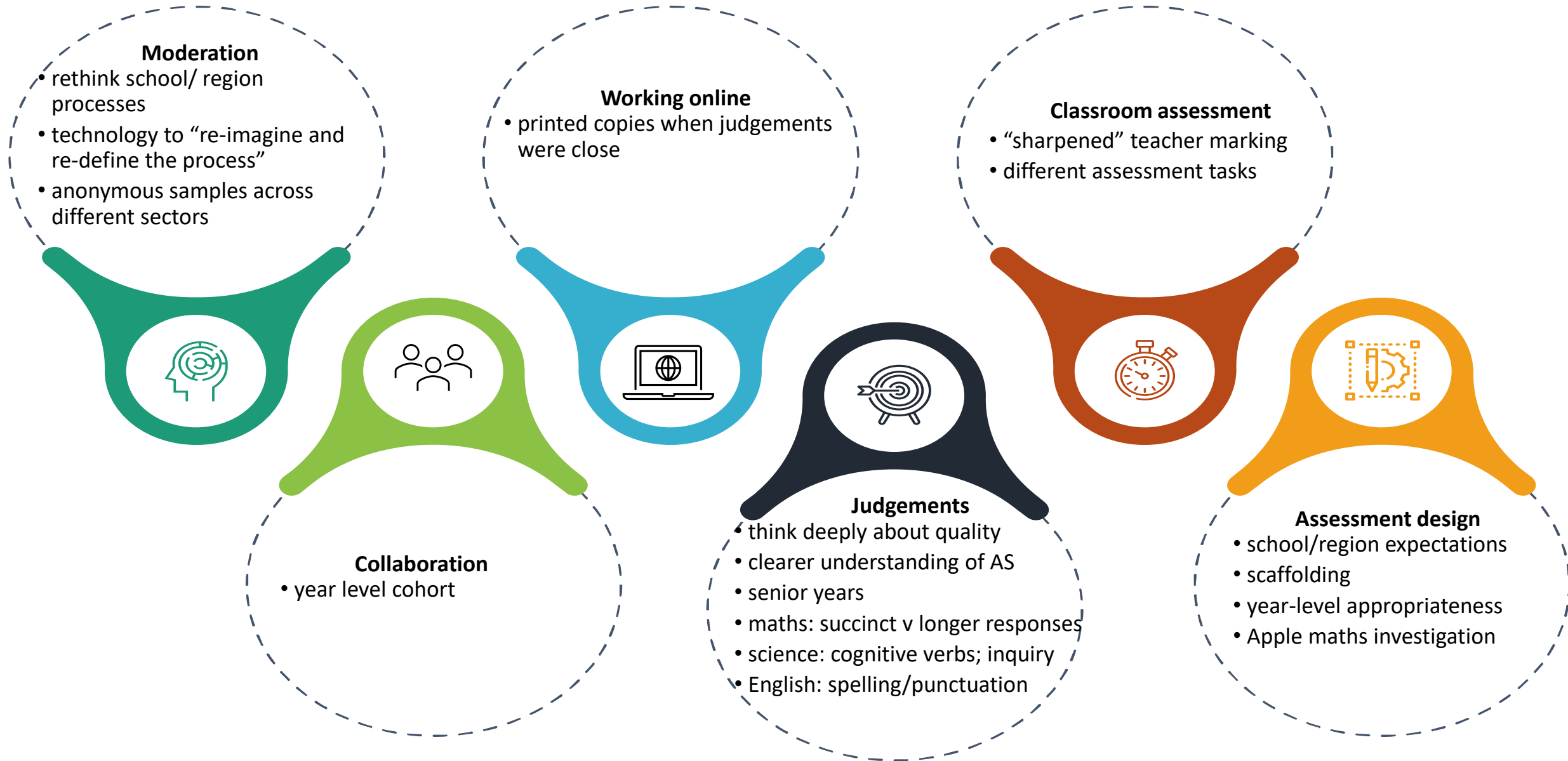
1. Median scale location for A level is only somewhat higher in Year 6 compared to Year 4
2. Relatively large range of scores selected for Year 4 D level
3. Somewhat larger variation for Year 8 B level, with a negatively skewed distribution

Stage 3 Standard Setting: Findings science



1. Median scale location for A level standard is higher in Year 6 compared to Year 4
2. Relatively large range of scores selected for Year 4 C level
3. Outliers present across year levels
4. Significant variation for Year 6 D level

Stage 3 Standard Setting: Participant Feedback



Sample Information					
Subject					
Strand/Topic					
Task					
Year/Class		Date			
Teacher Name					
Overall grade	A	B	C	D	E
Summary justification					
Strengths of the performance					
Areas for improvement					
Next steps for teaching					

Stage 4 Development of cognitive commentaries

Sample Information					
Subject	Science				
Strand/Topic	Chemical sciences investigation				
Task	Investigating the properties affecting the use of ochre				
Year/Class	4X	Date 18/11/2021			
Teacher Name					
Overall grade	A	B	C	D	E
Summary justification					
The student has demonstrated a mix of scientific understanding with general knowledge and communicated mainly through informal everyday language.					
Strengths of the performance					
The student has demonstrated that they can complete a sentence to describe their prediction and reason for their prediction. Within their attempt to craft an investigation question, they have identified the variable they are measuring. They have identified potential risks in terms of the materials used during the experiment. Throughout the response, they have referred to the variable they are collecting data on, i.e., 'smudge'.					
The student has used different terms to describe the properties of materials, such as, smooth, runny, sticky, bumpy, lighter, gooey. In the analysis, they have referred to some of those terms and used the word, 'properties'. They've referred to their common understanding of some of these materials and on the impact this will have. For example, they've drawn on their experience of working with glue and their knowledge that glue will dry hard to predict it will be the best material in terms of smudge resistance.					
Areas for improvement					
The student requires further support to craft an investigation question and identify a method for data collection using provided scaffolds.					
The recording of observations should be structured using a similar sequence for each test to show commonalities and make comparisons. For example, in this task, observations could be described in terms of stickiness, texture, ability to move material down the page. Quantitative data could also be incorporated using diagrams showing where the smudge is and a measurement. Diagrams have been					
Next steps for teaching					
Focus will be on developing scientific language through building on everyday descriptive words (e.g., 'runny') to more scientific words (e.g., 'fluid', 'viscosity'). Word walls will be used to show the build from everyday to scientific words.					
Scaffolding will be provided to support the student to think further about different aspects of questions. For example, asking about the properties of the observed materials followed by how those properties relate to use.					



Sample Information					
Subject	Science				
Strand/Topic	Chemical sciences investigation				
Task	Investigating the properties affecting the use of ash				
Year/Class	JK	Year	14/15/2021		
Teacher Name	Kerina/Lenora				
Overall grade					
	A	B	C	D	E
Summary justification					
The student has demonstrated a use of scientific understanding with general knowledge and conceptualisation through effective scientific reasoning.					
Strengths of the performance					
The student has demonstrated that they can consistently demonstrate their problem-solving and provide a reason for their problem. Whilst their attempt to craft an investigation question, they have identified the variables they are measuring. They have identified potential risks to some of the materials and their conduct during the experiment. Throughout the response, they have referred to the variables that they are collecting through the 'measured'.					
They have used different terms to describe the properties of materials, such as smooth, shiny, brittle, strong, hard, light, green in the samples. They have indicated aspects of their process and used the word 'temperature'. They've referred to their common understanding of some of these materials and the reason they will break. For example, they've stated that their experience of working with glass and their knowledge that glass will shatter is evident it will be the best material in terms of strength.					
Areas for improvement					
The student requires further support to craft an investigation question and identify a method for data collection using practical activities.					
The recording of observations should be structured using a table response for each task to show comparisons and make comparisons. For example, in this task, observations could be described in terms of observations, features, ability to meet material down the page. Student's data could also be investigated using diagrams showing when the change is at a measurement. Diagrams have been used very well in the response. Tables could also be used, for example, from 0 to 100 to show a 5 percent amount of change.					

Sample Information					
Subject					
Strand/Topic					
Task					
Year/Class		Date			
Teacher Name					
Overall grade					
	A	B	C	D	E
Summary justification					
Strengths of the performance					
Areas for improvement					
Next steps for teaching					

Copyright Cambridge University



ARC Linkage project (EP18010004): Improving teacher assessment capabilities using exemplars and online moderation

GUIDE TO PREPARING A COGNITIVE COMMENTARY

Welcome to Stage 4 of the Online Moderation project. Your time is appreciated, and we hope that your experience in this project will be a valuable professional learning opportunity.

During Stage 3 teachers identified samples that best represented a standard (A-E) based on provided descriptors. Teachers drew on their evaluative experience and expertise to make an on-balance decision, taking into consideration the standard descriptor and allowing for discretion to weight different criteria.

Part A of Stage 4 provides participants with the opportunity to document the thinking that occurs in this judgement process, that is, create a cognitive commentary to support the sample's awarded grade. After viewing instructional videos and sample cognitive commentaries, participants will use a template to write a cognitive commentary for each of their allocated 2-3 samples. Teacher feedback on the process of constructing the commentary and possible uses will be sought via a Reflection Form. The TRS (where applicable for Stage 4a) is one full day, this is a generous estimate of time required for the task.

In **Part B**, the submitted commentaries will be merged to produce a composite text that will be shared in small group meetings led by the Project Team. In these meetings points of difference will be discussed to reach agreement on the final commentary that participants believe reflects the assigned standard. Teacher feedback on the process of constructing the commentary and possible uses will be sought during these meetings or in post-meeting reflections.

The final commentary and work samples will be exemplars used in Stage 5 to trial their value in supporting teachers to make judgements of student work at the A-E levels for years 4, 6, and 8 in English, science and maths.

This document includes an overview of a cognitive commentary. The instructional videos, exemplars, templates, and A-E standard descriptors are provided on the project website; access details are on page 2. **The password is in the covering email.**

Cognitive commentary

A cognitive commentary is an explanation of how an overall judgement of the quality of student work is reached; it identifies the elements that came to the forefront in the appraisal process, taking into consideration the strengths and limitations of the work to arrive at an overall judgement. Next-step teaching decisions based on the characteristics of the work can also be included. The commentary is attached to the work sample and used by teachers to anchor the judgement to a standard of quality. It can be shared with colleagues in discussions about assessments or in moderation meetings, with parents when discussing student progress, and with students so that they develop understandings about quality and how to progress their learning.

The commentary addresses the following **key questions**:

1. What knowledge and skills are being demonstrated in the performance?
2. What are the areas for improvement on the performance?
3. Based on your assessment, what is the on-balance or overall grade you would award?
 - a. What comparisons have been applied in reaching the overall grade?
 - b. How have the strengths and weaknesses in the performance been combined to reach an overall grade?
4. What would be the next-step teaching required to progress learning for this student?
 - a. What resources could be drawn on to support this learning?

Resources



Professor Claire Wyatt-Smith
 Director, Institute for Learning Sciences & Numerical Education, Australian Catholic University

Stage 4 Development of Cognitive commentaries: Participants

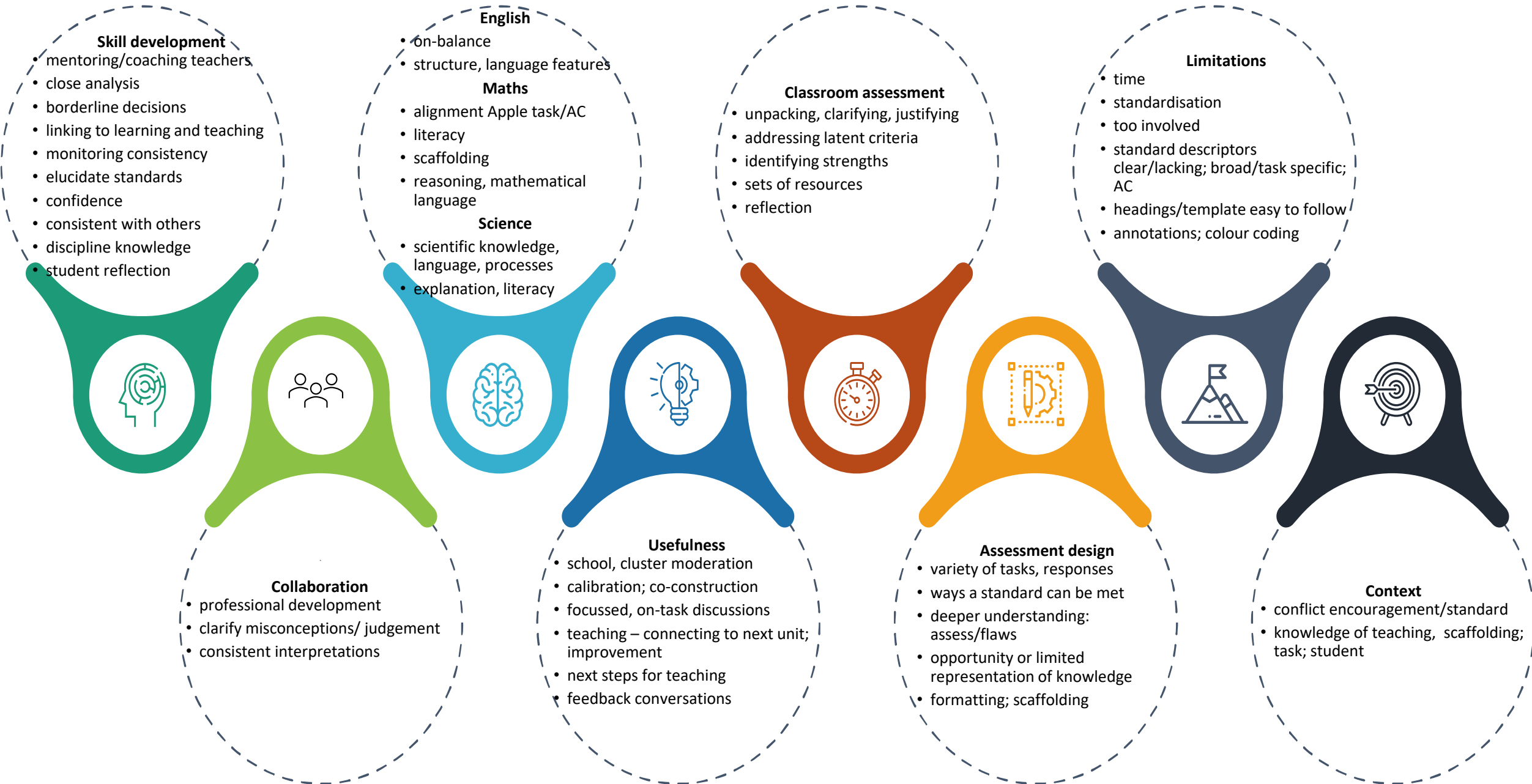
Region	No. of participating schools			
	English	Maths	Science	Total
Central Queensland	0	2	2	4
Darling Downs South West	1	2	1	3*
Far North Queensland	0	1	1	2
North Queensland	1	1	0	2
Metropolitan	3	3	5	8*
North Coast	0	1	0	1
South East	1	1	3	3*
Total	6	11	12	23

*some schools participated in more than one discipline

Discipline	Year Level	Teachers	DoE PPOs	Total Participants	Cog Comms written	Reflection Forms rec'd	
						T'chr	PPO
English	Yr4	3	2*	5	15	3	
	Yr6	1	5*	6	13	1	
	Yr8	4	3*	7	17	4	
Maths	Yr4	6	2*	8	18	6	
	Yr6	11	4*	15	30	11	
	Yr8	3	2	5	15	3	
Science	Yr4	5	2*	7	16	5	
	Yr6	11	2*	13	28	11	
	Yr8	2	2*	4	10	2	
Total		46	11	56*	162	46	10*

*some PPOs participated in more than one year level submitting one discipline combined reflection form

Stage 4 Development of Cognitive commentaries: Participant Feedback



Analysis of cognitive commentaries

Suggested improvements:

- core or extended knowledge and skills
- general or specific foci

Alignment between areas of improvement and next steps teaching strategies

Within strategies:

- what was to be taught/improved
- how it was to be taught

Mean number of improvement points per CC

Grade	English	Maths	Science
A	2.67	1.83	3.75
B	4	2.62	4.92
C	4.5	5	5.91
D	3.78	4.5	4.81
E	5	3.75	4.44

Mean number of next teaching steps per CC

Grade	English	Maths	Science
A	3	1.92	3.13
B	3.3	1.85	3
C	4.4	3.21	3
D	3.78	3.25	4
E	3.68	3.08	3.55



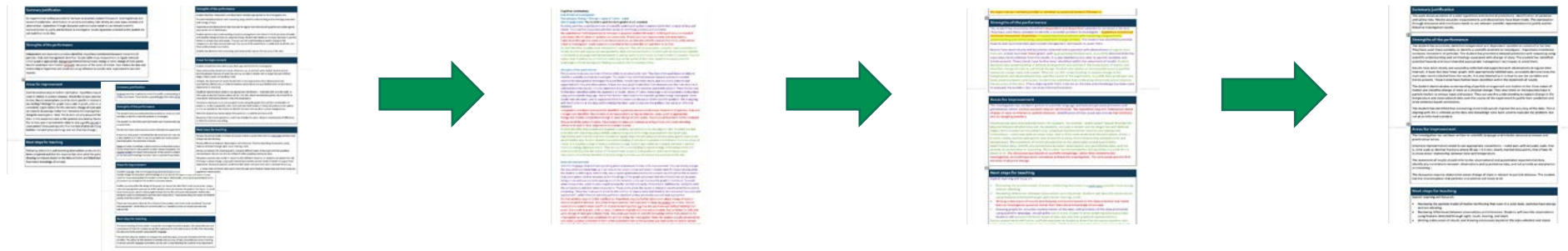
Areas for improvement primarily focused on basic skills of the subject areas



Next steps teaching strategies were often generic rather than specific to the response

e.g., scaffold, provide opportunities for..., utilise high impact strategies

Stage 4 Cognitive commentaries: Meetings



Discipline	Year Level	Teachers	No. of meetings
English	Yr4	3	1
	Yr6	1	1
	Yr8	4	1
Maths	Yr4	6	2
	Yr6	9	3
	Yr8	2	1
Science	Yr4	4	1
	Yr6	10	3
	Yr8	1	1
Total		40	14

Region	No. of participating schools			
	English	Maths	Science	Total
Central Qld	0	2	2	4
DDSW	1	1	1	3
Far North Qld	0	1	1	2
North Qld	1	1	0	2
Metro	3	3	4	7*
North Coast	0	0	0	0
South East	1	1	1	2*
Total	6	9	9	20

Outcome

- endorsement of 39 exemplars with cognitive commentaries
- remaining 6 samples
 - 4 samples endorsed with adjusted A-E standards
 - 2 assessment tasks deemed unsuitable and withdrawn

Based on teacher endorsements, no exemplar with cognitive commentary was provided for:

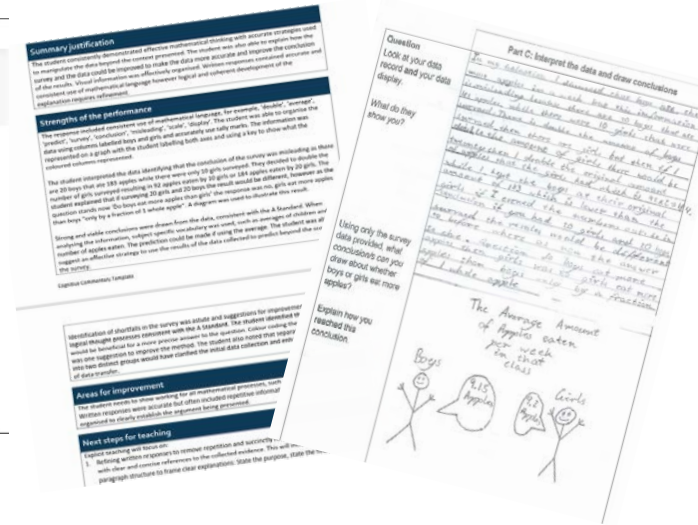
- English Yr4 E; Yr8 B
- Maths Yr8 A, E
- Science Yr4 A

Stage 5

Trial of resources

Discipline	Year Level	Teachers
English	Yr4	7
	Yr6	8
	Yr8	10
Maths	Yr4	5
	Yr6	5
	Yr8	4
Science	Yr4	3
	Yr6	7
	Yr8	7
Total		56

Number of tasks moderated			
English	Maths	Science	Total
124	65	85	274

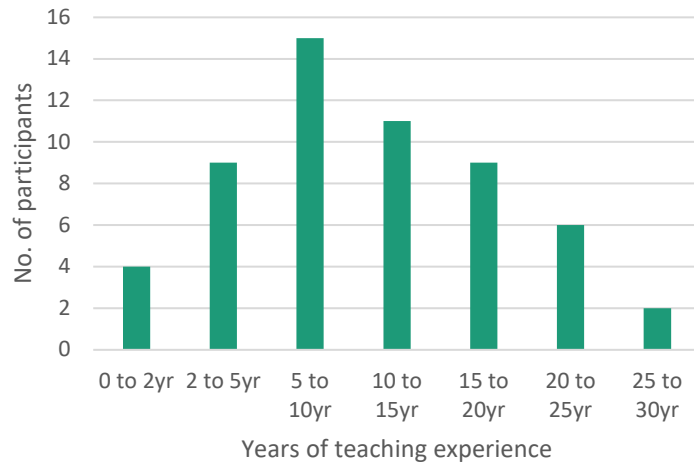


ARC LP18010046: Improving teacher assessment capability using scaled annotated exemplars of achievement standards in online moderation

MODERATION FORM (Group 1)

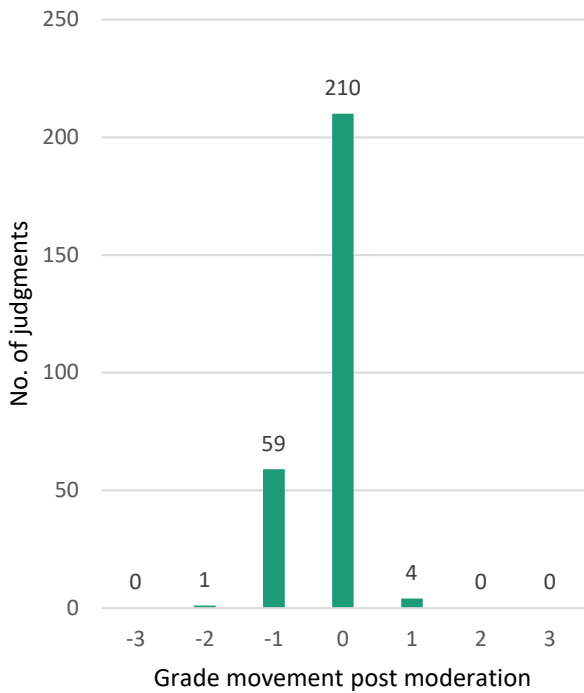
Name	4	6	8
Subject	English	Maths	Science

Sample Number	Original Grade	Moderated Grade
1		
2		
3		
4		
5		



2 sets (A-E) per year level/discipline expert review

Stage 5 Trial
of resources:
Overall
results



Teacher moderated grade							
Teacher original grade	A	B	C	D	E	Original total	% Changes
A	45	13	0	0	0	58	22
B	0	44	16	0	0	60	27
C	0	1	62	19	1	83	25
D	0	0	2	42	11	55	24
E	0	0	0	1	17	18	6
Moderated total	45	58	80	62	29	274	23

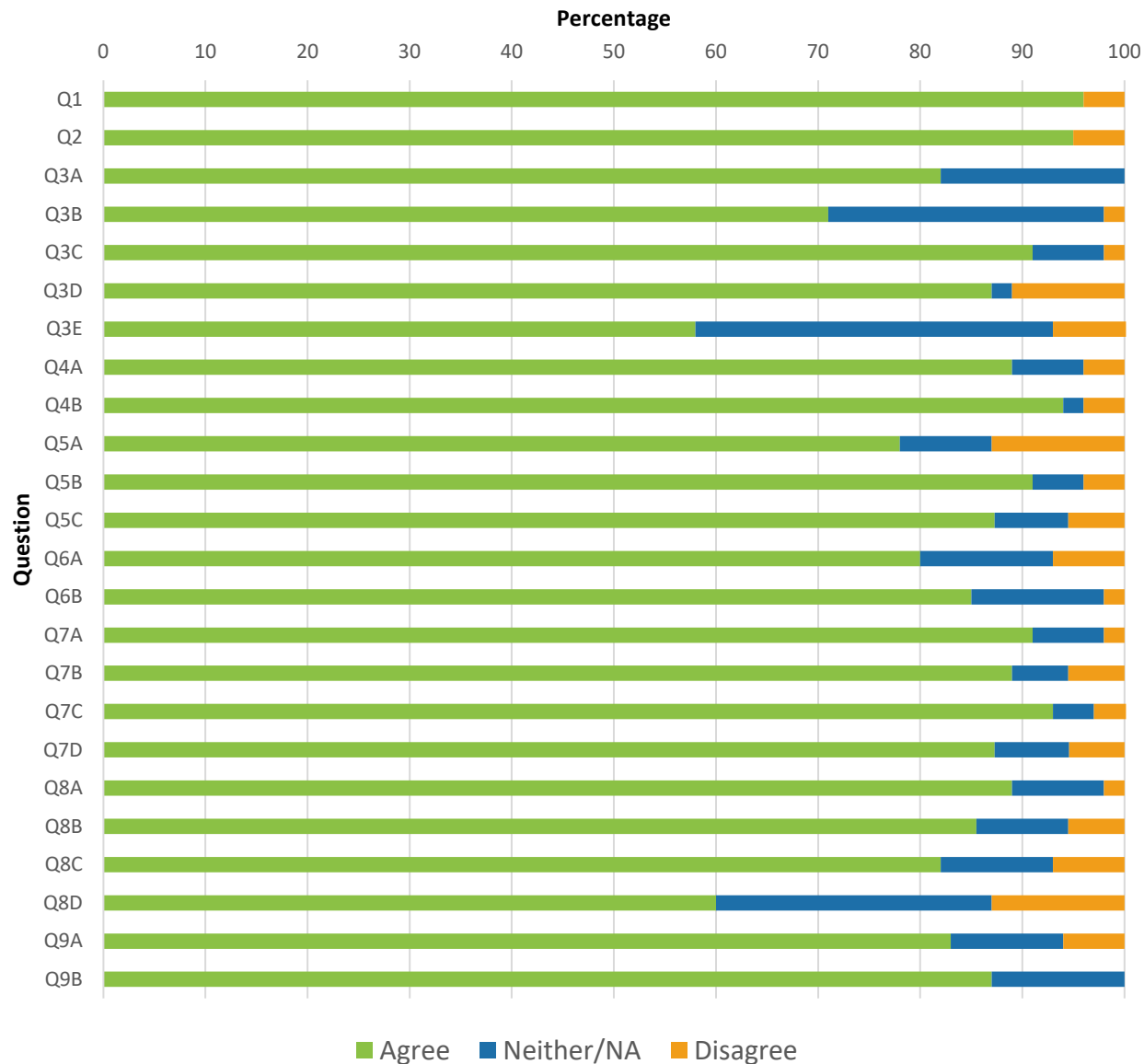
Stage 5: Tracking grades across stages

Discipline/ Yr Level	Stage 1 teacher grade matches endorsed grade		Stage 1 grade and endorsed grade within one grade level	
	No.	%	No.	%
English	8/15	53	15/15	100
Maths	8/13	62	13/13	100
Science	4/15	27	12/15	80
Yr4	7/15	47	13/15	87
Yr6	8/15	53	15/15	100
Yr8	5/13	38	12/13	93
Overall	20/43	47	40/43	93

* Two Yr8 Maths tasks were withdrawn

Discipline	Year Level	Grade: Stage 1	Grade: Stage 3	Grade: Stage 4
English	4	A	B	B
		B	A	A
		C	C	C
		D	D	D
		E	E	E
	6	A	A	A
		B	B	B
		C	C	C
		D	D	D
		E	E	E
	8	A	A	A
		B	B	B
		C	C	C
		D	D	D
		E	E	E
Maths	4	A	A	A
		B	B	B
		C	C	C
		D	D	D
		E	E	E
	6	A	A	A
		B-	B	B
		C	C	C
		D	D	D
		E	E	E
	8	A	A	A
		B	B	B
		C	C	C
		D	D	D
		E	E	E
Science	4	A	A	B
		A	B	C
		A	C	C
		C-	D	D
		D	E	E
	6	A	A	A
		A-	B	B
		B	C	C
		D	D	D
		D+	E	E
	8	A	A	A
		A	B	B
		A	C	C
		D	D	D
		D+	E	E

Stage 5: Teacher Reflection Summary



Q1. Website easy to access and use

Q2. Exemplars viewed without delay

Q3. A exemplar useful in moderating judgements

Q3. B exemplar useful in moderating judgements

Q3. C exemplar useful in moderating judgements

Q3. D exemplar useful in moderating judgements

Q3. E exemplar useful in moderating judgements

Q4A. Exemplars assisted me to better understand standards

Q4B. Exemplars assisted me to identify evidence of standards

Q5A. Would use exemplars when planning to align teaching to standards

Q5B. Would use exemplars when planning to demonstrate knowledge and skills required

Q5C. Would use exemplars when planning to improve unit/assessment items

Q6A. Would use exemplars when teaching to explain task/standard requirements

Q6B. Would use exemplars when teaching to remind me of teaching strategies

Q7A. Would use exemplars to calibrate before grading

Q7B. Would use exemplars when making judgements

Q7C. Would use exemplars during moderation

Q7D. Would use exemplars to guide student feedback

Q8A. Would use exemplars to identify common areas of strengths/weaknesses

Q8B. Would use exemplars to assist with next-step teaching

Q8C. Would use exemplars to evaluate effectiveness of teaching strategies

Q8D. Would use exemplars in parent meetings

Q9A. Would write cog comms for selected samples to carry forward my thinking

Q9B. Would write cog comms for selected samples to model the standard for other teachers

Stage 5 Trial of resources: Participant feedback

Moderation

- positive, non-threatening approach
- easy to see comparison between my work and grade boundary
- confirm or rethink grades for my own skills
- capture thinking about judgement for moderation meeting



Teaching & Learning

- knowledge of content/standards
- common misconceptions/errors
- various levels
- ideas for teaching strategies
- new to year level, out-of-field, next year teachers
- analysis, reflection
- robust discussion of performance
- co-creating own sets
- differentiation



Classroom assessment

- improve task, marking guide
- consistency; remove latent criteria
- understanding/confirming/justifying grade
- examples of judgement decisions
- borderline judgements
- distinction D/E standards
- different representations of same quality



Improvements

- alignment AC, EQ marking guides, task, current school practices
- scaffolding esp. maths
- more exemplars available
- identification of threshold/mid-range samples
- adding annotations/descriptors to exemplar
- tasks: same/different; quality
- time allocation



Parent meetings & Reporting

- detailed justification
- clear, understandable breakdown of strengths and weaknesses
- goal setting
- next steps for teaching



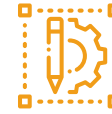
Planning

- see students' thinking
- clarify task expectations
- introducing new/refined unit
- distinguish differences in levels
- possible teaching strategies



Students

- visible expectations; authentic models of standards
- 'bump up' performance
- comparisons when explaining results/difference between standards
- feedback; goal setting
- strategies: think-alouds, guided deconstruction, student grading, learning wall



Teacher voices

no matter where you are in your career... we actually found 'These **questions** are not as rich as the others, how can **we improve** that?' And so, only through actually **analysing the student's responses** and then by recognising the **test let them down**...but we've now **modified that assessment piece**

it was **easy to see** the comparative links between my work and the grade boundary points to confirm or make me **rethink grade points for my own skills**

But now through this process I feel like I have **authority** behind me to say I've done four stages of this, I am able to moderate different levels of work, and I also, **believe in myself** more.

this is a great way that we can **establish moderation** in our school ... It helped us, it was the fact that we could **actually identify the strengths of the tasks** was how we made a **discerning judgment between the A and B**

...we have **so many new staff**, I'm currently **upskilling** four staff ...I see this as being so valuable because it would take away so much of that additional upskilling time... our new staff don't have the knowledge

I have also **trialled using these** ... with the result that **students were more confident and informed** when writing their statements in T2 compared to T1. Their **writing was more concise, and they improved their results...**

focusses the teacher on the evidence and its alignment to standards, and uses real student work as exemplars and descriptions to compare with

*we're redoing our **parent-teacher interviews** ... saying 'What are the **strengths** ... areas of **improvement**, and our **next steps for teaching**'. So, conveying to the parents what we actually see next steps **curriculum wise***

Outcomes

- a **transparently ordered set of A-E exemplars with cognitive commentaries** to support sustainable and consistent judgement decisions
- an innovative **approach to moderation using online technologies** suited to **national engagement**
- evidence of **how achievement standards are applied to student work**, demonstrating how aspects of criteria are combined in different ways in the representation of a standard
- focus areas for building teachers' assessment capabilities and in **teacher preparation**
- analysis of teacher talk showed investment in Queensland assessment and moderation processes to **ensure fair and equitable assessment** for students
- how exemplars, illustrative of the required standard, together with evaluative commentaries of the judgement decision **increased self-confidence** in judgement decisions and promoted **shared understandings** of quality within teaching teams
- cognitive commentaries functioned to stimulate discussions of **next-step teaching** and how teaching, learning and assessment can **align** in practice.

Future Study Design to Investigate State-wide Judgement Consistency

- larger study
- participating teachers selected using a *multi-stage stratified random sampling design*
- stratified by geographical regions or location type
- further within-region strata: ICSEA categories (low, medium, high), school size/type, year level

Sampling of schools could take place within each of these defined strata to ensure that differences in judgement consistency by teachers can be examined between these categories of schools.



Project outputs

1. Exemplars and associated resources Online moderation tool and resources (OMTAR)
2. Reports and Policy briefs – informing future directions – Implications for policy, practice and research going forward
3. Promotional videos
4. Conference presentations
 - a. Adie, L. *Assessment moderation: Is it fit-for-purpose?* Australian Council for Educational Research (ACER) virtual conference (Monday 22 August to Thursday 25 August 2022). <https://vimeo.com/acerorg/review/745707290/0eb1fdf427>
 - b. Wyatt-Smith, C., & Adie, L. *Developing and using cognitive commentaries of assessment decisions with a focus on next step teaching and progressing student learning*. Modules (including videos) developed for Catholic Schools NSW (10 October 2022)
5. Journal articles

Humphry, S., & Bredemeyer, K. (2022). Pairwise comparison scale extension using core linking sets. *Frontiers in Education*, 7:826742. doi:10.3389/feduc.2022.826742

In production:

 - a. Improving teachers' ability to use assessment data to identify next learning steps: The potential of structured analysis. *British Educational Research Journal*.
 - b. Evidence in practice: Teacher assessment work and identity. *Teaching and Teacher Education*.
 - c. An institutional ethnography of teachers' assessment and moderation practices. *Australian Educational Researcher*.
 - d. Using pairwise comparison and ordered exemplars as a basis for standard setting of narrative writing. *Frontiers in Education*.
 - e. Teacher judgement of student work: The infrastructure for building dependability. *British Educational Research Journal*.

Big findings: policy, practice and research going forward. Where to?

