ofqual

Grading of Vocational and Technical Assessments

Mary Ward House
8d0da34dfe

slido.com
#ofqual

Welcome and housekeeping

Roger Taylor

Conference outline

1. Policy context
   - Phil Beach

3. Practitioners' experiences
   - Carina Fagan
   - Barry Smith
   - Adam Sturt
   - Chair Phil Beach

5. Going forward
   - Tim Oates

2. Taking stock – the grading research
   - Paul Newton

4. Scholars' reflections
   - Steve Higgins
   - Gerard Lum
   - Matt Homer
   - Chair Beth Black

6. Paul Newton
   - Conclusions and close

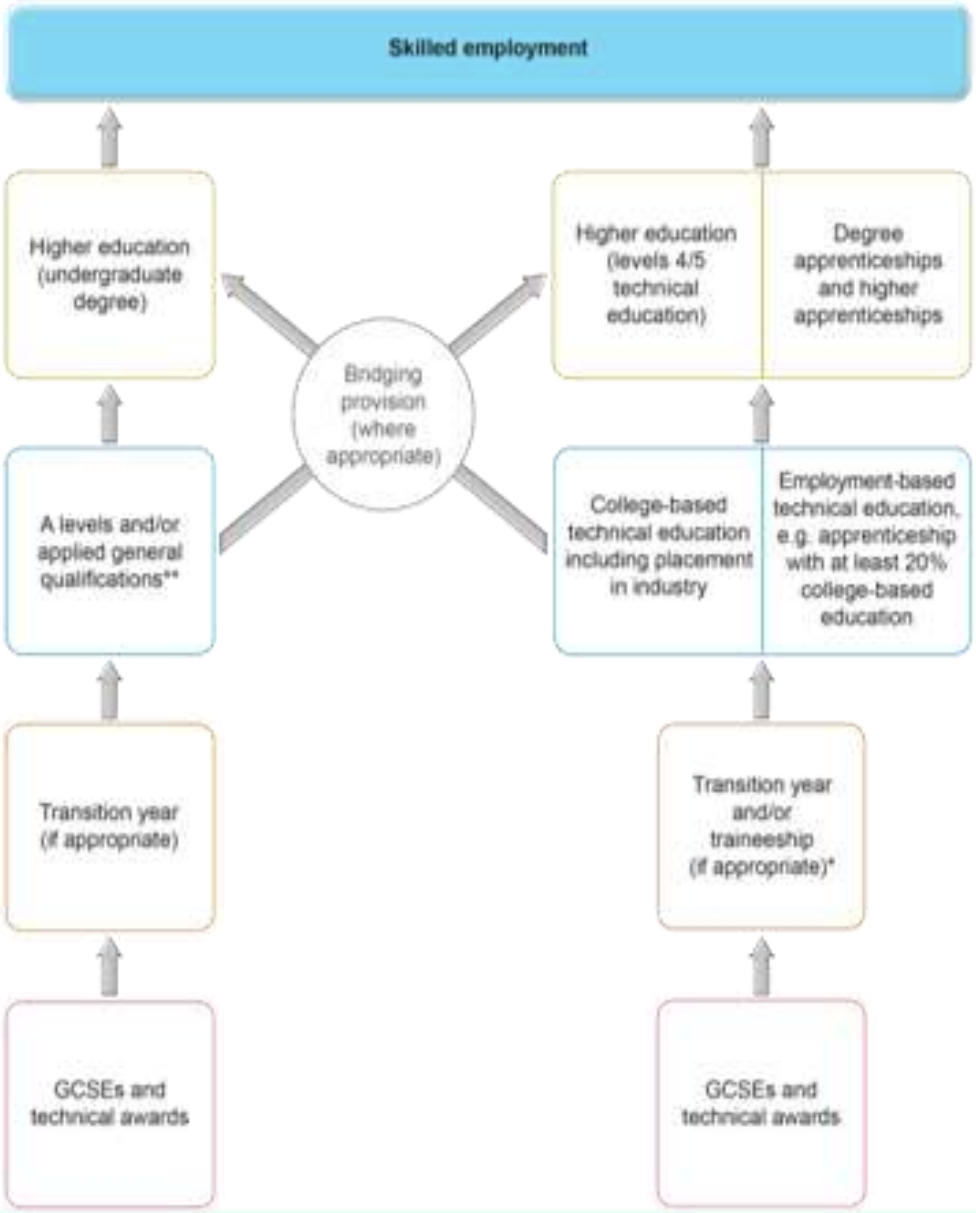
Policy context

Phil Beach

Overview of vocational qualification policy changes

- Over the past decade, we've seen several major reviews and inquiries.
- A key theme across all reviews was about increasing the currency of vocational and technical qualifications.
- Government policy responses have included:
  - requiring grading structures within performance table qualifications;
  - review of funding formulas;
  - introduction of graded end-point assessment;;
  - increasing focus on ‘parity’ between general and vocational qualifications.

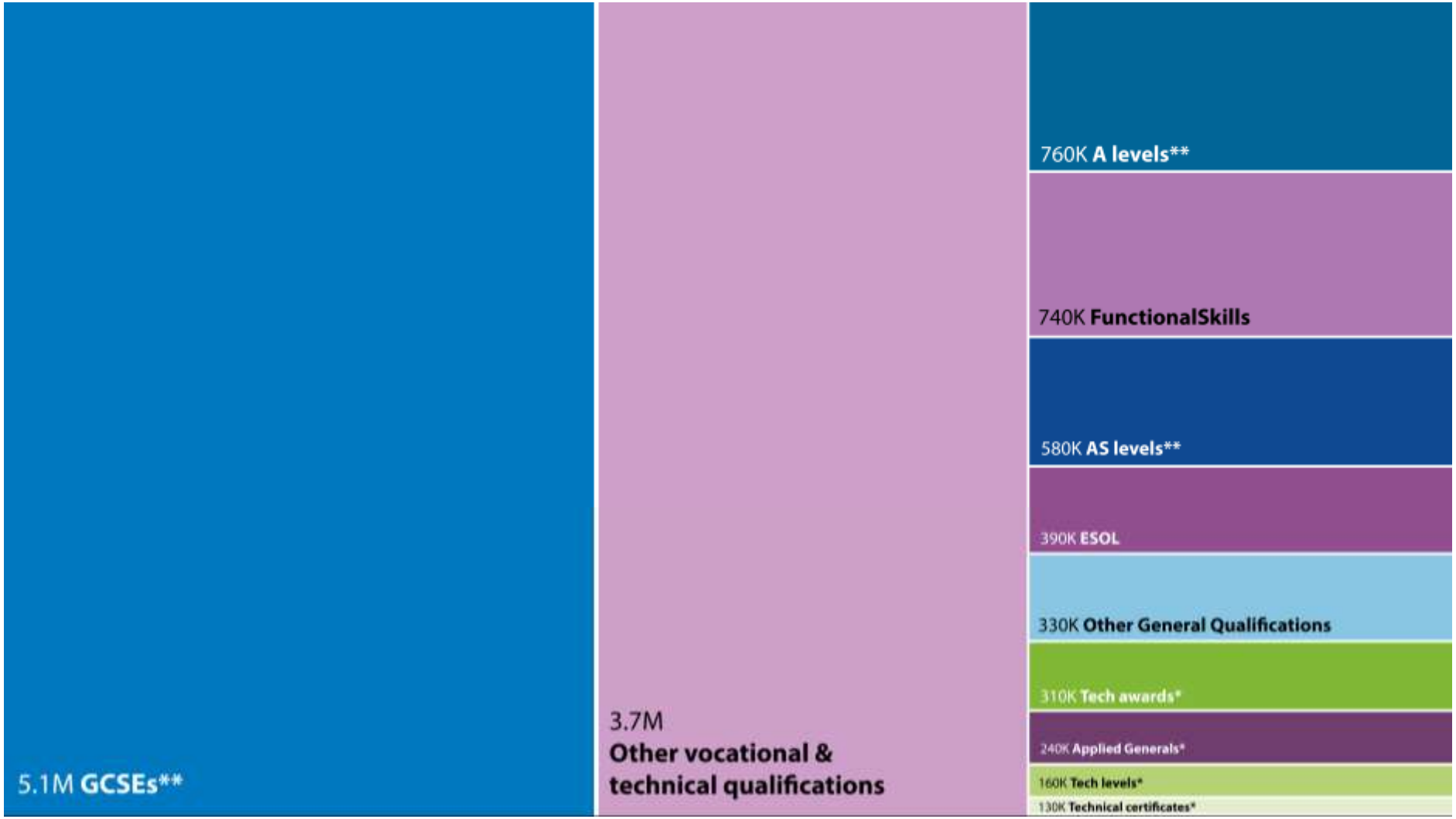
T levels and skills plan



Certificates issued, year to Q1 2018
Numbers rounded

*Included in 2017-2020 performance tables
**Based on 2016-17 academic year

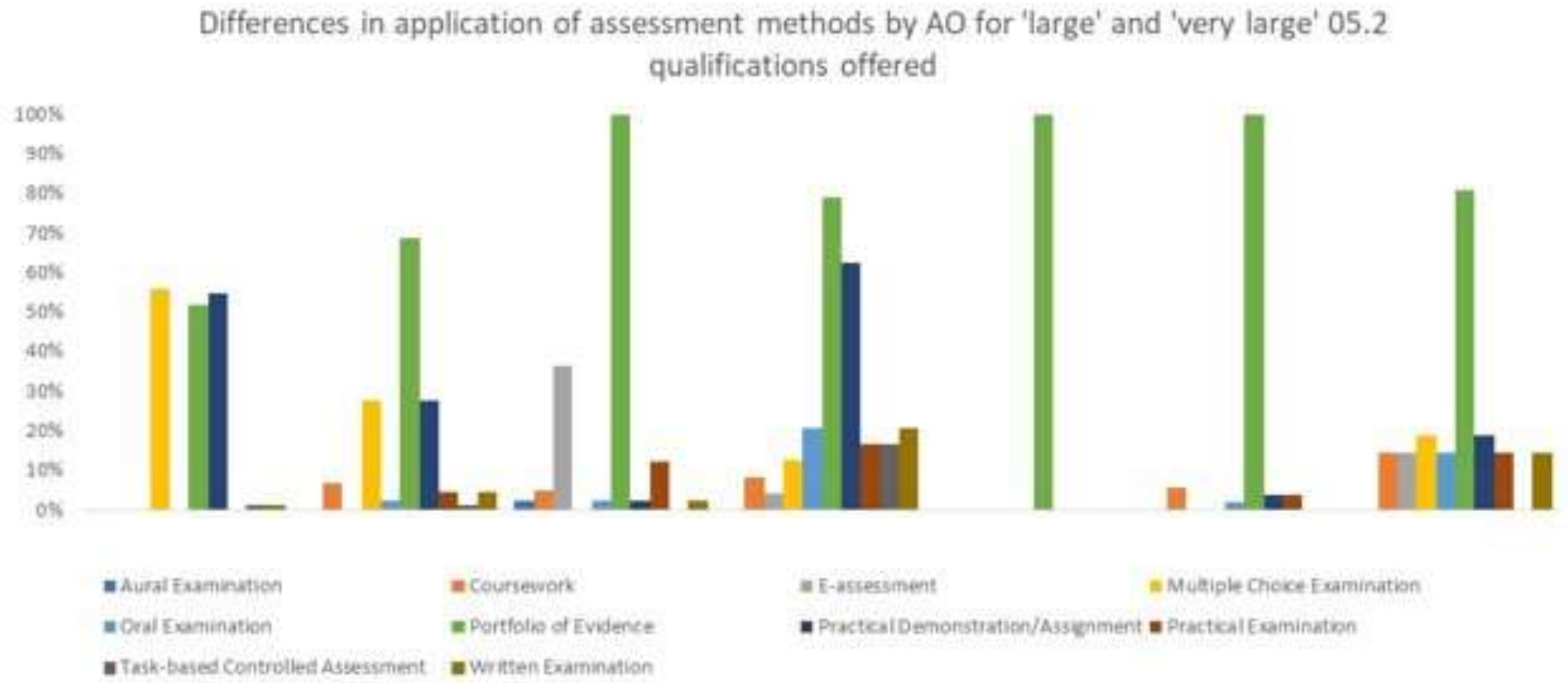
ofqual



Qualifications available - July 2018
Numbers rounded

*Included in 2017-2020 performance tables

SSA5.2 Buildings and Construction as an example of assessment methods



Ofqual's approach

- We've increased our commitment to researching vocational and technical qualifications
- We want to improve understanding of best practice in assessment and grading within the sector
- There will be choices to be made on assessment and grading across vocational qualifications

Grading Vocational & Technical Assessments

Paul E. Newton

11 December 2018

- The Grade Debate
- Grade Expectations
- Make England Grade Again

Grading Vocational & Technical Assessments

Grading Vocational & Technical Assessments

- Applied Generals
- Graded Examinations
- ESOL examinations
- End-Point Assessments
- Technical Qualification examinations
- etc.

We’re launching a conversation on grading…

Credit: Jesús Alenda

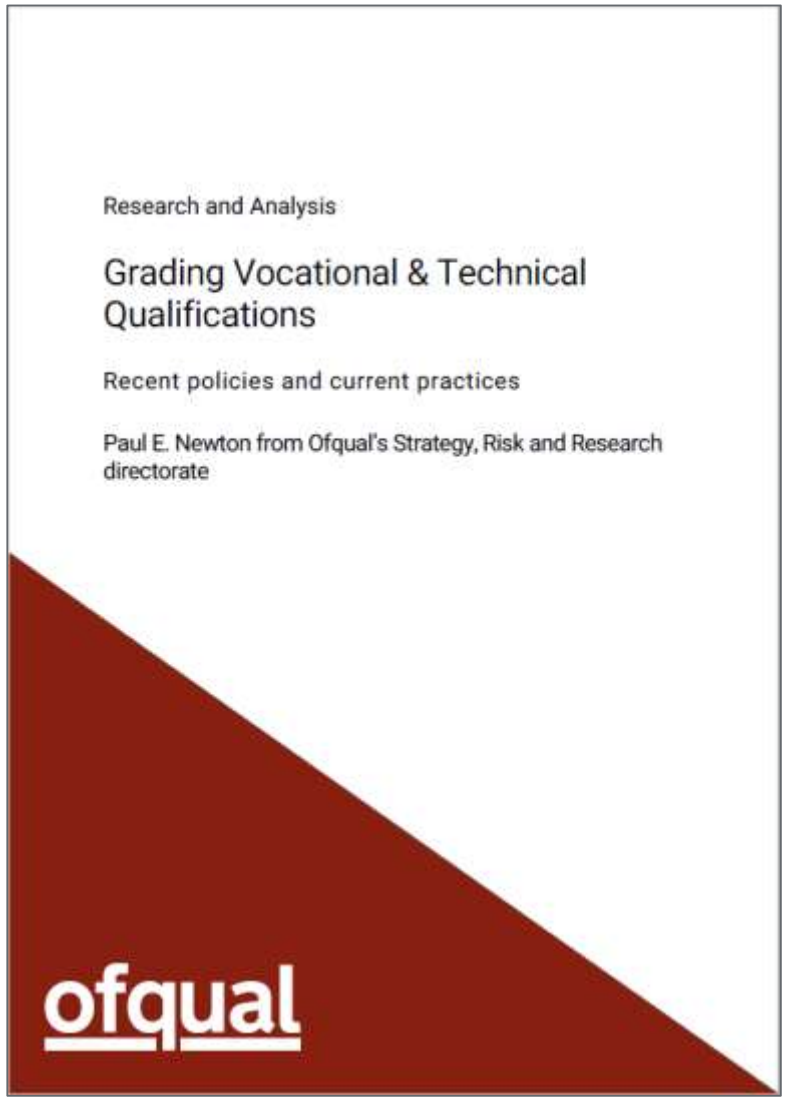
- on the basis of a limited amount of information on the ‘how and why’ of grading practices
- with the intention of generating far more information and far deeper understanding

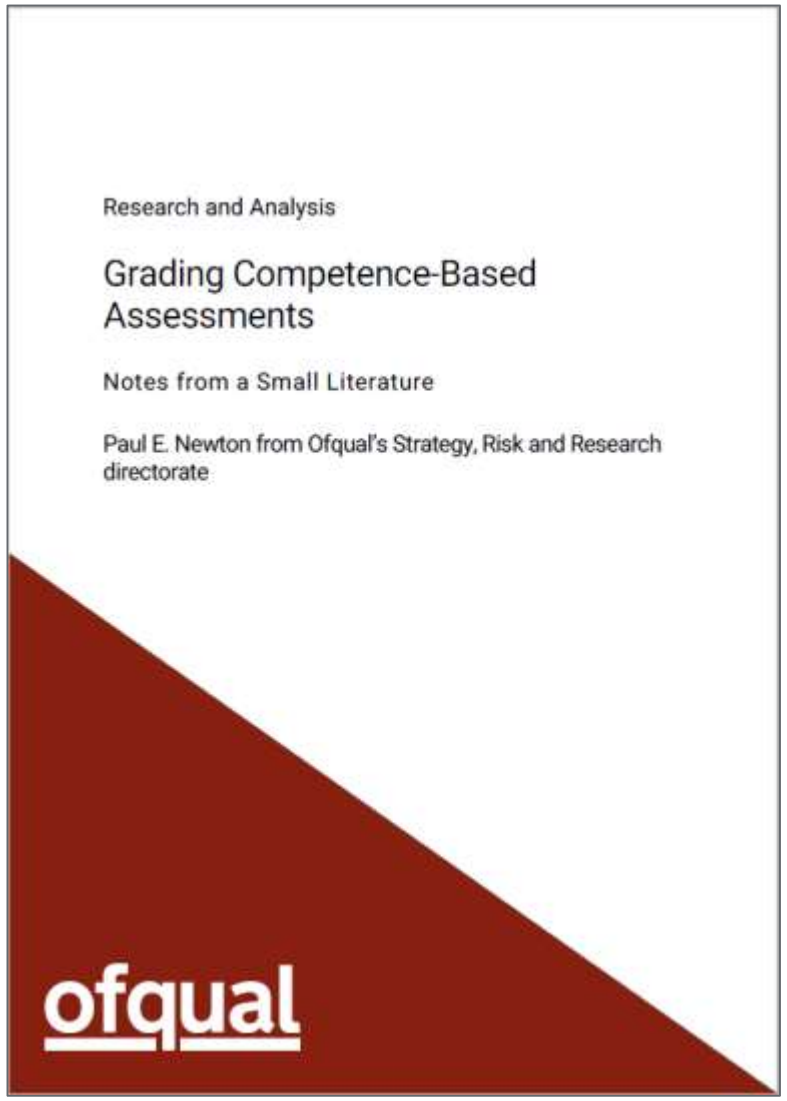
Why are we (Ofqual) initiating a conversation on grading?

- Vocational & Technical Assessments tend to be a bit different
- The increasing use of grading makes them increasingly different
- The literature on grading VTAs (like ours) is remarkably small
- Grading practices in England appear to be remarkably divergent

- Why are different qualifications graded in different ways?
- Are certain grading practices more valid than others?
- What are the pros and cons of alternative practices?

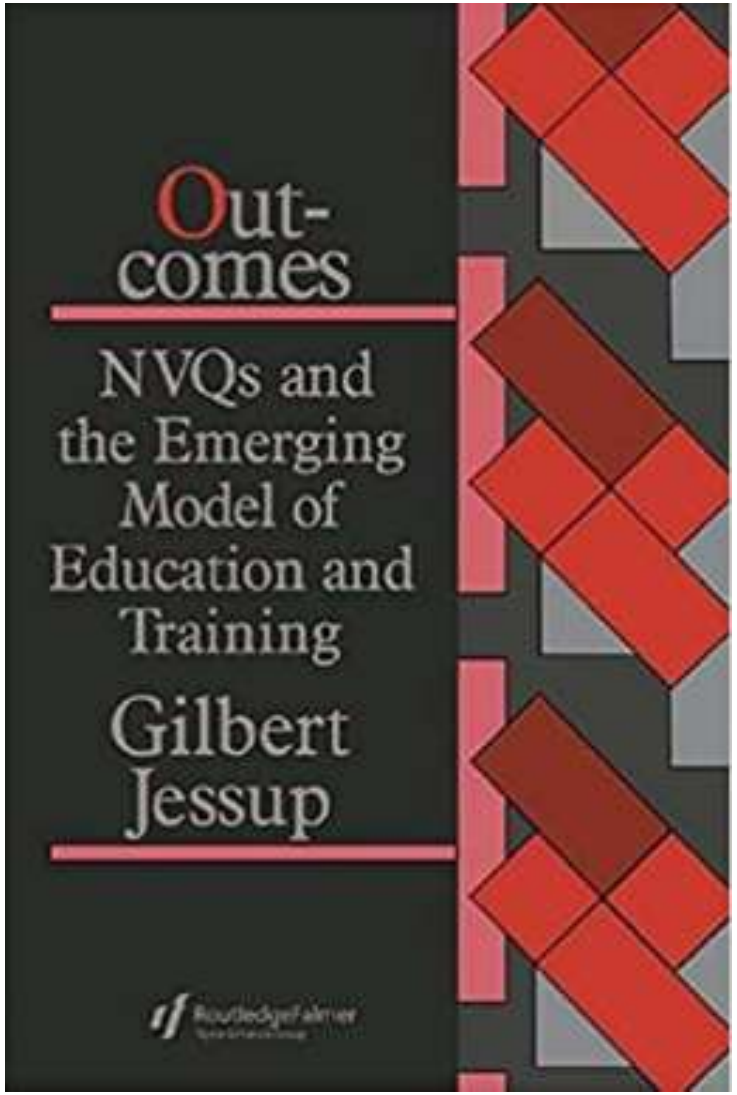
A new set of information on grading VTAs





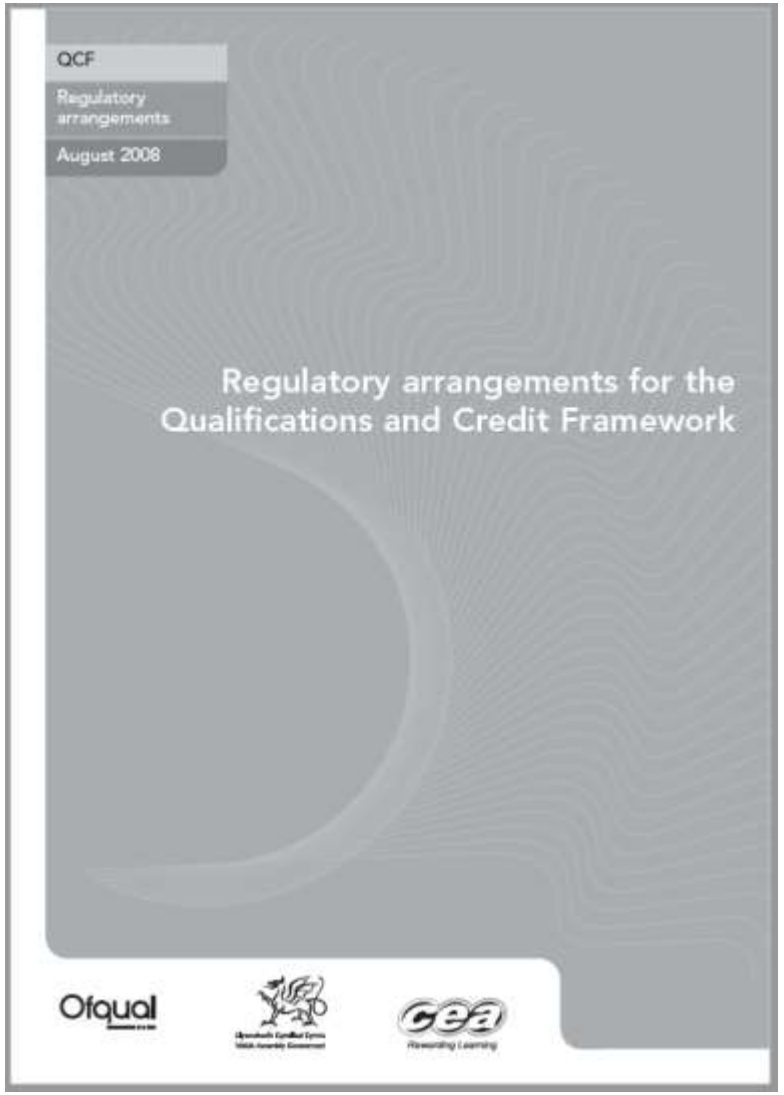
- **What today ISN’T about**
  - **Critiquing policy**
  - **Compliance**
  - **Criticising**
- **What today IS about**
  - **Understanding practice**
  - **Research & analysis**
  - **Collaborating**

Why do VTAs in England tend to be a bit different?



Competence-Based Assessment

- **Competence**
  - associated with professional or occupational standards
- **Atomistic** specification
  - multiple units
  - multiple learning outcomes
  - multiple assessment criteria
- **Mastery** aggregation



Qualifications and Credit Framework

- **Majority of VTQs ended up being recognised under the QCF**
  - **11,291 (64%) of Ofqual-regulated qualifications classified as ex-QCF as at 28/11/2017**
- **CBA-like design requirements**
  - **Atomistic specification**
  - **Mastery aggregation**

Policy drivers versus practical constraints

Professor Alison Wolf,
Baroness Wolf of Dulwich

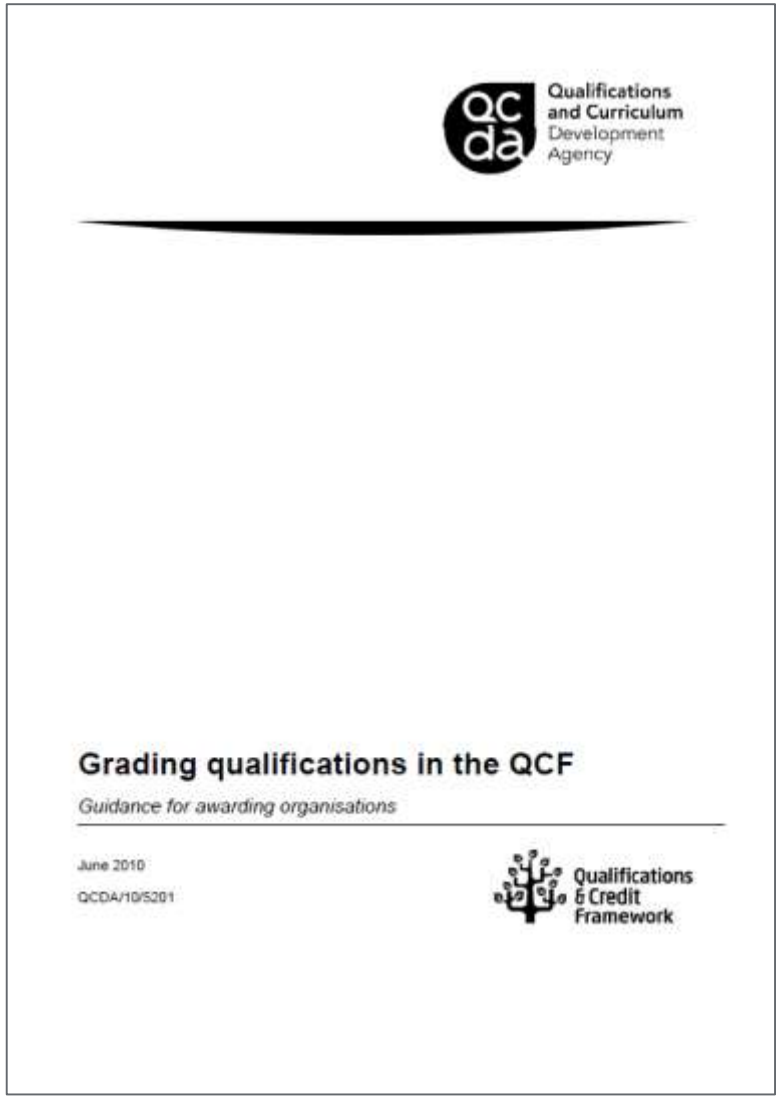
"One fundamental problem for educational institutions is that they are competence-based awards which are meant to attest that someone has reached a particular threshold or level of workplace competence. […] However, any candidate for educational progression needs to demonstrate not only a specific level of competence, but relative performance, otherwise the qualification is of little use to selectors. While a considerable number of QCF awards do allow for grading, awarding bodies told the review that this had been very difficult to achieve."

(Wolf, 2011, pp.86-7)

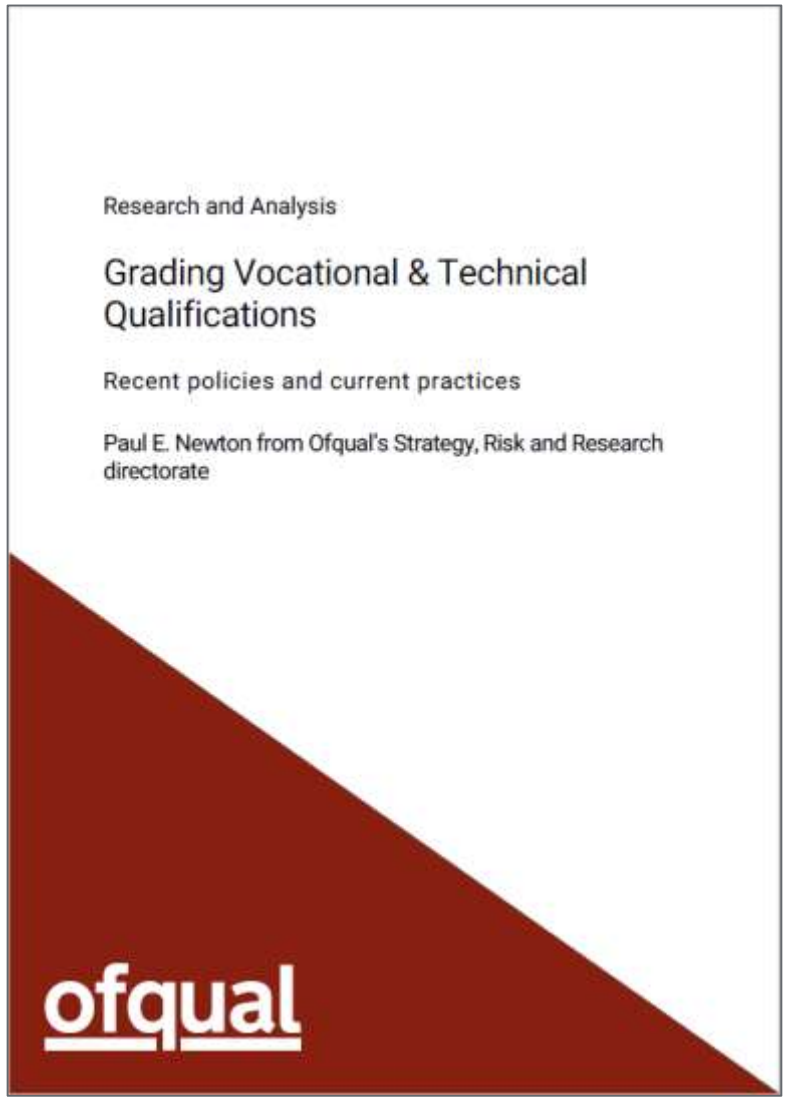
Difficult but not impossible

Regulatory Arrangements (1.28 & 1.29)

- Criteria for a pass must be:
  - based exclusively on the assessment criteria in the unit or units on which the component or qualification is based
- Criteria for higher grades must:
  - build explicitly on the assessment criteria of […] units within the […] qualification
  - be consistent with the overall level of the qualification […]
  - be clearly distinguished from achievements at the next level of the QCF



A limited amount of information on grading VTQs (including CBVQs)



As at 28/11/2017

- **15,852 qualifications classified as something other than General Qualifications**
  - **13,685 awarded pass only**
  - **2,167 awarded higher grades too**
    - **2,021 'classic' structure (ie M, D, D*)**
    - **146 alternative structure (eg A to E)**

Largest clusters

Grading Structure	EPA	ESOL	Entry	FSMQ	FSQ	Higher	Key Sk.	NVQ	Occnl	Pr. Le.	Project	QCF	VCE	VRQ
P/M		6									46			5
P/D	4					1			1		3			10
P/M/D		54	3			48		1	11		609			465
P/M/D/D*									1		250			261
Double P/M/D														36
Double P/M/D/D*											77			57
Triple P/M/D														1
Triple P/M/D/D*											45			26

Available qualifications, as at 28/11/2017.

Small-scale exploratory survey

- **Opportunity sample** from the 10 (structure x type) cross-classifications that contained more than 40 available qualifications.
- Aimed to identify a sample of qualifications that seemed **likely to reflect a variety** of grading approaches.
- Qualifications were only included if sufficient information could be **located on the internet** concerning their approach to grading.

Caveat emptor

- Only 18 qualifications
- Only 15 awarding organisations

Qual. abbreviation	Qual. grade	Qual. type	No. units	Unit level(s)	External assess.	Internal assess.	Unit/Ass. grading	All U/A graded?
RSPH L4 Food Award	P/D	QCF	(2 exams)	4	100%		P/D	Y
NOCN L1 ESOL Cert.B2	P/M/D	ESOL	4	1	100%		P/M/D	Y
BTEC L4 Policing HNC	P/M/D	HIGHER	8	4		100%	P/M/D	Y
BTEC L5 Electrical HND	P/M/D	QCF	~ 16	4 & 5		100%	P/M/D	Y
TCL L3 Speech Grade 8	P/M/D	QCF	(1 exam)	3	100%		P/M/D	Y
UWLQ L2 Oral Grade 4	P/M/D	QCF	(1 exam)	2	100%		P/M/D	Y
ABC L4 Art F.Dipl.	P/M/D	QCF	8	3 & 4		100%	P/M/D	N
ATHE L3 Business Dipl.	P/M/D	QCF	4	3		100%	P/M/D	Y
Cskills L2 Carpentry Dipl.	P/M/D	VRQ	8	1, 2 & 3	8 units	5 units	P/M/D	N
CIBTAC L3 Beauty Dipl.	P/M/D	VRQ	6	3	100%		P/M/D	Y
IMI L3 Vehicle Dipl.	P/M/D	VRQ	10	2 & 3	8 units	10 units	P/M/D	N
NCFE L2 Bus.Skills Dipl.	P/M/D/D*	QCF	6	2		100%	P/M/D	Y
RSL L3 Music Dipl.	P/M/D/D*	QCF	~ 11	3		100%	P/M/D	Y
VTCT L3 Barbering E.Dipl.	P/M/D/D*	VRQ	10	3 (mainly)	6 units	10 units	P/M/D	N
AQA L3 App.Bus. Cert.	P/M/D/D*	VRQ	3	3	2 units	1 unit	P/M/D	Y
OCR L3 Business T.Dipl.	P/M/D/D* (x2)	QCF	12	3		100%	P/M/D	Y
OCR L3 IT T.Dipl.	P/M/D/D* (x2)	VRQ	11	3	3 units	8 units	P/M/D	Y
BTEC L3 Dental E.Dipl.	P/M/D/D* (x3)	QCF	16	3		100%	P/M/D	Y

Striking variety of approaches to grading

- Current practice in grading VTAs in England isn't underpinned by a straightforward, generally accepted, set of principles governing good practice.
- That doesn't imply that any of the observed grading approaches reflects anything less than good practice.
- A fairer conclusion is that we will need to undertake more work with AOs to understand whether or not we could establish a set of principles to underpin grading practices for VTAs in England.

Some findings

- with a particular focus on Competence-Based Vocational Qualifications (cf. Graded Exams, ESOL, etc.)
- recalling that ‘caveat emptor’ clause

Level 3 Diploma for Music Practitioners (comprising 11 units)

	Assessment Criteria (Pass Criteria)
AC 1.1	Explain the marketing strategy for an agreed product including a range of physical and digital promotional devices
AC 1.2	Implement a promotional campaign, via appropriate promotional methodology (e.g. The Promotion Mix), for an agreed musical product or service
AC 1.3	Produce a Promotion E-Portfolio for the campaign in 1.2 that identifies the marketplace and appropriate audience for the product and includes examples of promotional materials
AC 1.4	Evaluate learning from the unit by creating a career focused action plan that highlights how they might apply the promotion techniques learned to their own products in the future

- **'Music Promotion' unit**
  - **single LO**
    - **understand/apply concept of music promotion**
  - **four AC**
  - **different task for each AC**
  - **assessed by teacher**
  - **must satisfy all four AC to pass the unit**

Common themes (despite divergent practices)

Approaches to specifying measurement standards

- cf. CBA model
  - atomistic specification of measurement standards

Approaches to aggregating measurement information

- cf. CBA model
  - mastery measurement principle

Theme 1: specifying measurement standards

- How do CBVQs specify standards for higher grades?
- 'Typically' by adding two extra layers of criteria.

		Pass Criteria (AC)	Merit Criteria	Distinction Criteria
LO1	AC 1	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
	AC 2	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
	AC 3	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
LO2	AC 1	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
	AC 2	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
LO3	AC 1	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
	AC 2	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah

L3 Diploma for Music Practitioners

Performance *quality*

	Pass Criteria (AC)	Merit Criteria	Distinction Criteria
AC 1.1	Explain the marketing strategy…	Explain **with clarity and detail** the marketing strategy…	**Comprehensively** explain the marketing strategy…
AC 1.2	Implement a promotional campaign…	Implement a **clear and detailed** promotional campaign…	Implement a **rigorous** promotional campaign…
AC 1.3	Produce a Promotion E-Portfolio…	Produce a **clear and detailed** Promotion E-Portfolio…	Produce a **comprehensive** Promotion E-Portfolio…
AC 1.4	Evaluate […] by creating a career focused action plan…	Evaluate […] by creating a **clear and detailed** career focused action plan…	Evaluate […] by creating a **perceptive and rigorous** career focused action plan…

Criteria for higher grades must:

- **build explicitly on the assessment criteria of […] units within the […] qualification**
- **be consistent with the overall level of the qualification […]**

L3 Dental Technology Diploma

Performance *complexity*

	Pass Criteria (AC)	Merit Criteria	Distinction Criteria
2	**Explain** the importance of the chemical, biological, and mechanical properties in the selection of materials used in dental technology	**Discuss** the key points relating to chemical, biological, physical and mechanical properties of dental materials	**Analyse** the properties of a range of dental biomaterials, justifying their selection
3	**Explain** the use of dental waxes found within the dental laboratory	**Review** the properties of dental waxes	**Evaluate** waxes used in the dental laboratory
4	**Explain** the selection criteria applied to assess the suitability of using either gypsum or synthetic stone materials in the dental laboratory	**Discuss** the use of gypsum and synthetic stone materials used in the dental laboratory	**Critically evaluate** gypsum and synthetic stone materials used in the dental laboratory

Theme 2: aggregating measurement information

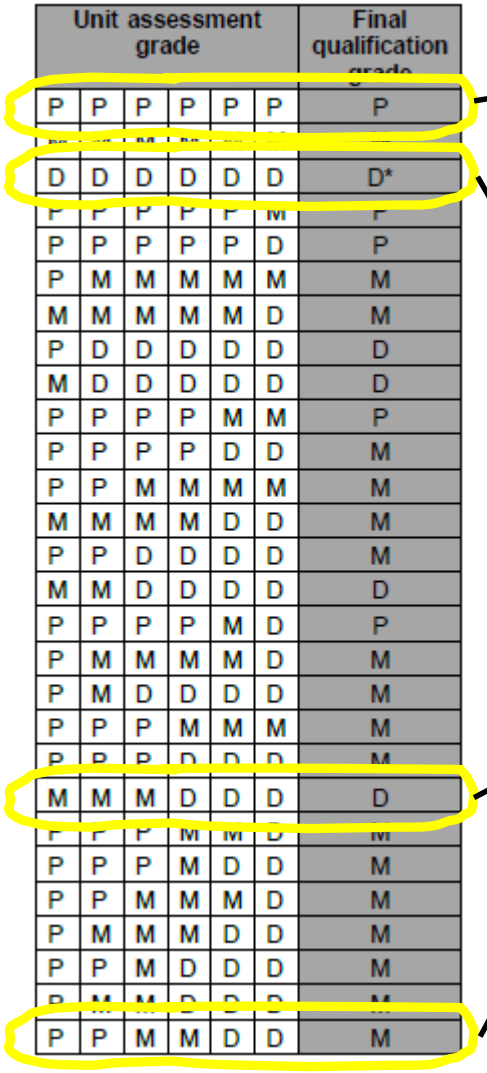
- How do CBVQs aggregate assessors' judgements for higher grades?
- 'Typically' in exactly the same way as for the passing grade, i.e. mastery.

		Pass Criteria (AC)	Merit Criteria	Distinction Criteria
LO1	AC 1	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
	AC 2	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
	AC 3	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
LO2	AC 1	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
	AC 2	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
LO3	AC 1	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah
	AC 2	Blah, Blah, Blah	Blah, Blah, Blah	Blah, Blah, Blah

L2 Skills for Business Diploma (within-unit aggregation)

Assessment Criteria		Pass Criteria [must satisfy all]	Merit Criteria [must satisfy all]	Distinction Criteria [must satisfy all]
1.1	**Identify the different types of document that could be used to present agreed information**	The candidate will […]		
1.2	**Assess the different formats that could be used within each of these documents**	The candidate will […]	The candidate will […]	
2.1	**Confirm the purpose, content, style, deadline for a range of products**	The candidate will […]	The candidate will […]	
2.2	**Obtain the required content for each document**	The candidate will […]	The candidate will […]	
2.3	**Assess the different types of technology that can be used to create each document**	The candidate will […]	The candidate will […]	
3.1	**Select the most appropriate format for each document**	The candidate will […]	The candidate will […]	
3.2	**Organise the structure and layout of their documents**	The candidate will […]	The candidate will […]	The candidate will […]
3.3	**Produce the documents**	The candidate will […]	The candidate will […]	The candidate will […]
3.4	**Correct any text or formatting errors**	The candidate will […]	The candidate will […]	The candidate will […]

L2 Skills for Business Diploma (across-unit aggregation)



Unit assessment grade						Final qualification grade
P	P	P	P	P	P	P
D	D	D	D	D	D	D*
P	P	P	P	P	D	P
P	M	M	M	M	M	M
M	M	M	M	M	D	M
P	D	D	D	D	D	D
M	D	D	D	D	D	D
P	P	P	P	M	M	P
P	P	P	P	D	D	M
P	P	M	M	M	M	M
M	M	M	M	D	D	M
P	P	D	D	D	D	M
M	M	D	D	D	D	D
P	P	P	P	M	D	P
P	M	M	M	M	D	M
P	M	D	D	D	D	M
P	P	P	M	M	M	M
M	M	M	D	D	D	D
P	P	P	M	D	D	M
P	P	M	M	M	D	M
P	M	M	M	D	D	M
P	P	M	D	D	D	M
P	P	M	M	D	D	M

P P P P P P = P

- **Mastery** principle

P P M M D D = M

- **Compensation** principle

M M M D D D = D

- **Charity** principle

D D D D D D = D*

- **Charity** principle (extreme)

L3 Applied Business Certificate

Performance Outcomes	Pass Criteria	Merit Criteria	Distinction Criteria
PO1 Investigate small business marketing communications	P1	M1	
	P2	M2	D1
PO2 Plan a marketing communications strategy	P3		
	P4	M3	D2
	P5	M4	
	P6	M5	D3
PO3 Develop a marketing communications mix	P7	M6	D4
	P8	M7	D5
PO4 Recommend a schedule of marketing communications	P9	M8	
	P10	M9	D6
Total no. criteria required for each unit grade	**10**	**15**	**20**

UMS conversion

- **40**
- **60**
- **80**

L3 Applied Business Certificate (across-unit aggregation)

Qualification grade	180 GLH (3 units) total: 300 UMS **UMS points**
Distinction*	270–300
Distinction	240–269
Merit	180–239
Pass	120–179

D* = more than 3 x D (3 x 80 = 240)

- Compensation principle

D = 3 x D unit points (3 x 80 = 240)

- Compensation principle

M = 3 x M unit points (3 x 60 = 180)

- Compensation principle

P = 3 x P unit points (3 x 40 = 120)

- Mastery principle

L3 Applied Business Certificate (within-unit aggregation)

Performance Outcomes	Pass Criteria	Merit Criteria	Distinction Criteria
PO1 Investigate small business marketing communications	P1	M1	
	P2	M2	D1
PO2 Plan a marketing communications strategy	P3		
	P4	M3	D2
	P5	M4	
	P6	M5	D3
PO3 Develop a marketing communications mix	P7	M6	D4
	P8	M7	D5
PO4 Recommend a schedule of marketing communications	P9	M8	
	P10	M9	D6
Total no. criteria required for each unit grade	**10**	**15**	**20**

You can fail an entire learning outcome and still pass the unit

The (CBVQ) atomistic approach to specifying standards isn't mandatory

Use the grading rubric below, to award the learner a grade for the synoptic assessment.

All or most of the characteristics must be achieved to award the given grade
Characteristics of a Pass
The learner carried out a complete barbering service which satisfied the essential requirements of the client and mostly aligned with the consultation. The learner produced a cut and finish, and trimmed/cut facial hair using suitable methods and techniques, including compliance with health and safety and product specifications. The finished look was suitable for the client. The learner communicated with the client to check expectations, maintain positive personal contact and conclude the service to the client's satisfaction. The learner completed the service within the scheduled time, whilst maintaining a clean and tidy work area. The learner demonstrated sufficient knowledge and understanding throughout the assessment. The learner explained some aspects of the service, with reasons for the decisions taken and made suggestions for improvement to personal performance.
Characteristics of a Merit
The learner systematically carried out the agreed services which fully aligned with the consultation and employer expectations. The learner produced a cut and finish of equal quality to achieve a pleasing overall result. The learner used a range of techniques proficiently which were suitably tailored for the client. The learner used communication skills to establish and maintain a positive relationship with the client throughout. The learner maintained customer service, client comfort and confidence. The learner managed time, materials and equipment effectively whilst maintaining a safe working area. The learner discussed methods and products for future hair care, including the services and products available from the business. The learner demonstrated awareness of environmental sustainability with some eco-friendly measures employed. The learner demonstrated a broad and detailed knowledge and understanding throughout the assessment. The learner justified most aspects of the service given relative to the client's needs and wants, with reasons for the decisions and actions taken. The learner identified key strengths, weaknesses and opportunities for improvement.
Characteristics of a Distinction
The learner systematically and seamlessly integrated services, which fully aligned with the consultation and exceeded employer expectations. The learner produced a result that demonstrated equal mastery in cut and finish, which delighted the client. The learner used a range of techniques with a high degree of skill, confidence, flow and fluidity that were suitably tailored for the client. The learner achieved a finish that demonstrated mastery, creativity and innovation. The learner used a communication strategy that was courteous, considerate and attuned to the clients responses, which established a positive relationship with the client throughout. The learner provided a high degree of customer service and actively sought opportunities to make the client feel special. The learner managed time, the work area, equipment and materials for optimal safety, efficiency and effectiveness. The learner used strategies to provide advice and guidance, and promote and sell products and services, discussing the relative advantages and disadvantages of future options; demonstrating a strong commercial focus. The learner demonstrated commitment to environmental sustainability employing eco-friendly measures throughout. The learner demonstrated a broad and comprehensive knowledge and understanding throughout the assessment. The learner reflected upon and justified all aspects of the service, covering the options available, and the reasons for the decisions and actions taken. The learner explained strengths and

L3 Barbering Extended Diploma

- **Terminal synoptic unit**
- **Externally set, internally marked**
- **Complete barbering service**

Holistic 'best-fit' grading judgement

"Assessors must not attempt to assign a grade to each and every task, instead they must weigh up the strengths and weaknesses of the service provided as a whole and decide on which grade best reflects the learner's overall performance."

Four different aggregatory principles

1. **mastery** – overall result represents (or tends towards) the **lowest** level of proficiency across a specified domain, or subdomain
2. **compensation** – overall result represents an **average** level of proficiency across a specified domain, or subdomain
3. **configuration** – overall result represents a particular **pattern**, or configuration, of proficiencies across a specified domain, or subdomain
4. **charity** – overall result represents (or tends towards) the **highest** level of proficiency across a specified domain, or subdomain

All of the sampled qualifications operated aspects of both mastery and compensation.

Many of the qualifications operated at least three of these principles simultaneously.

All sorts of technical issues arise – eg standardisation

"[…] the short history of NVQs has also been one in which the quest for clarity has produced an ever more complex and complicated 'methodology'. As with all competence-based systems, the assumption has always been that assessment will be unproblematic because it simply involves comparing behaviour with the transparent 'benchmark' of the performance criteria. Unfortunately, in practice this turns out not to be the case."

(Wolf, 1995, p.24)

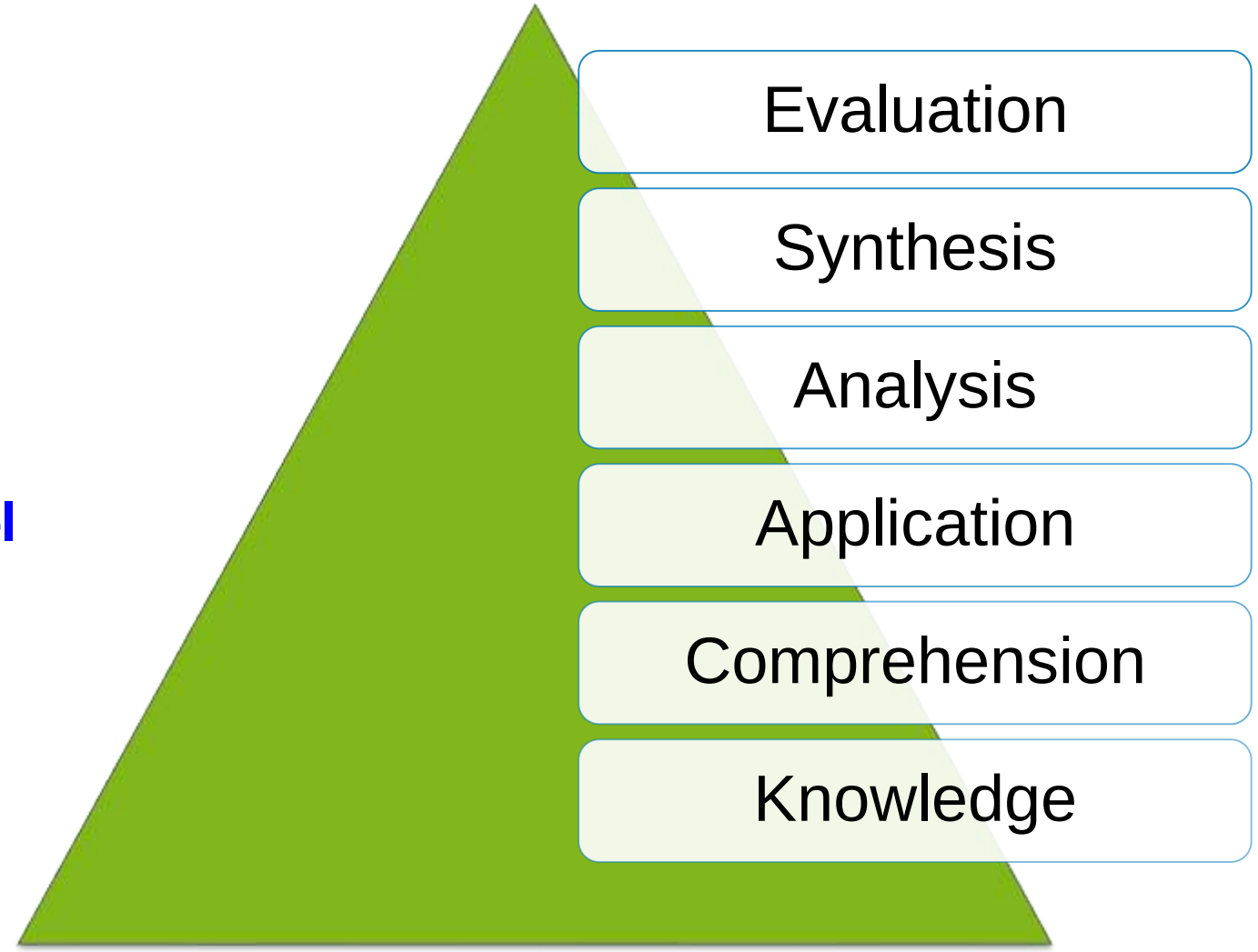
AC1.1

- explain the marketing strategy for an agreed product… (Pass)
- explain with clarity and detail the marketing strategy for an agreed product… (Merit)
- comprehensively explain the marketing strategy for an agreed product… (Distinction)

All sorts of technical issues arise – eg grading vs. levelling

Regulatory Arrangements

- **Criteria for higher grades must:**
  - **be consistent with the overall level of the qualification […]**
  - **be clearly distinguished from achievements at the next level of the QCF**



Level 2 Diploma in Skills for Business

Unit 10 Respond to change in a business environment

LO1 Understand the reasons for change in business

AC1.1 **State** why it is important for a business to change

AC1.2 **State** the risks associated with a business changing too quickly

AC1.3 **State** the risks associated with a business changing too slowly

LO2 Understand how change can affect people within a business

AC2.1 **Outline** positive effects change can have on people working in a business

AC2.2 **Outline** negative effects change can have on people working in a business

Level 3 Diploma in Skills for Business

Unit 10 Respond to change in a business environment

LO1 Understand change in business

AC1.1 **Explain** why it is important for a business to change

AC1.2 **Analyse** the positive and negative effects of change on a selected

AC1.3 **Compare** the risks of slow against rapid change within a business

AC1.4 **Compare** the benefits of slow against rapid change within a business

LO2 Understand how change can affect people within a business

AC2.1 **Explain** why people respond positively to change in a business

AC2.2 **Explain** why people respond negatively to change in a business

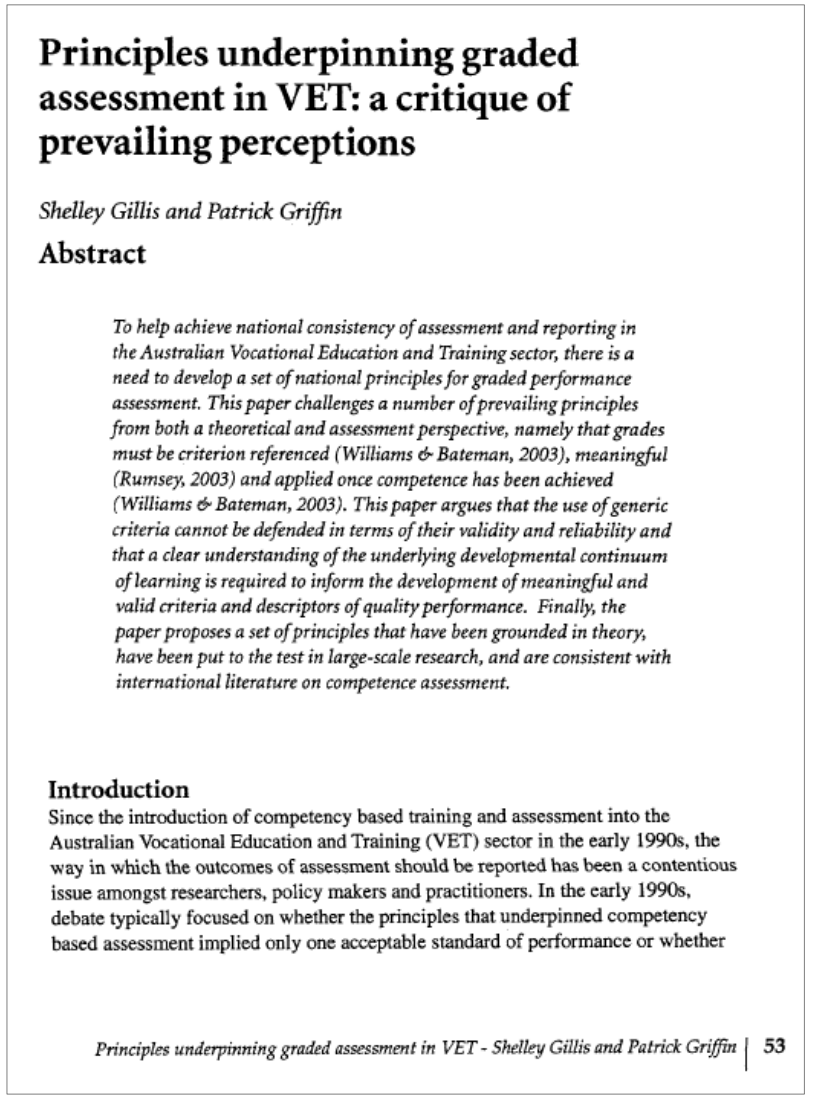
Unit 10	Pass	Merit	Distinction
L2 **AC1.1**	The candidate will **state** why it is important for a business to change	The candidate will **state** why it is important for a business to change, demonstrating **critical understanding**	[No D for this AC]
L3 **AC1.1**	The candidate will **explain** why it is important for a business to change	The candidate will **explain** **in detail** why it is important for a business to change	The candidate will give a **sophisticated explanation** of why it is important for a business to change

All sorts of conceptual questions arise

Is it good practice to use the same model for higher grades as for the passing grade?

Or should higher grades…

- be given a different kind of meaning
  - e.g. aptitude or diligence (vs. competence)
- be based on different criteria
  - e.g. generic (vs. competence-specific)
- involve a different aggregation process
  - e.g. compensatory (vs. mastery)
- be reported separately
  - e.g. Competent-with-Merit (vs. Merit)



Principles underpinning graded assessment in VET: a critique of prevailing perceptions

Shelley Gillis and Patrick Griffin

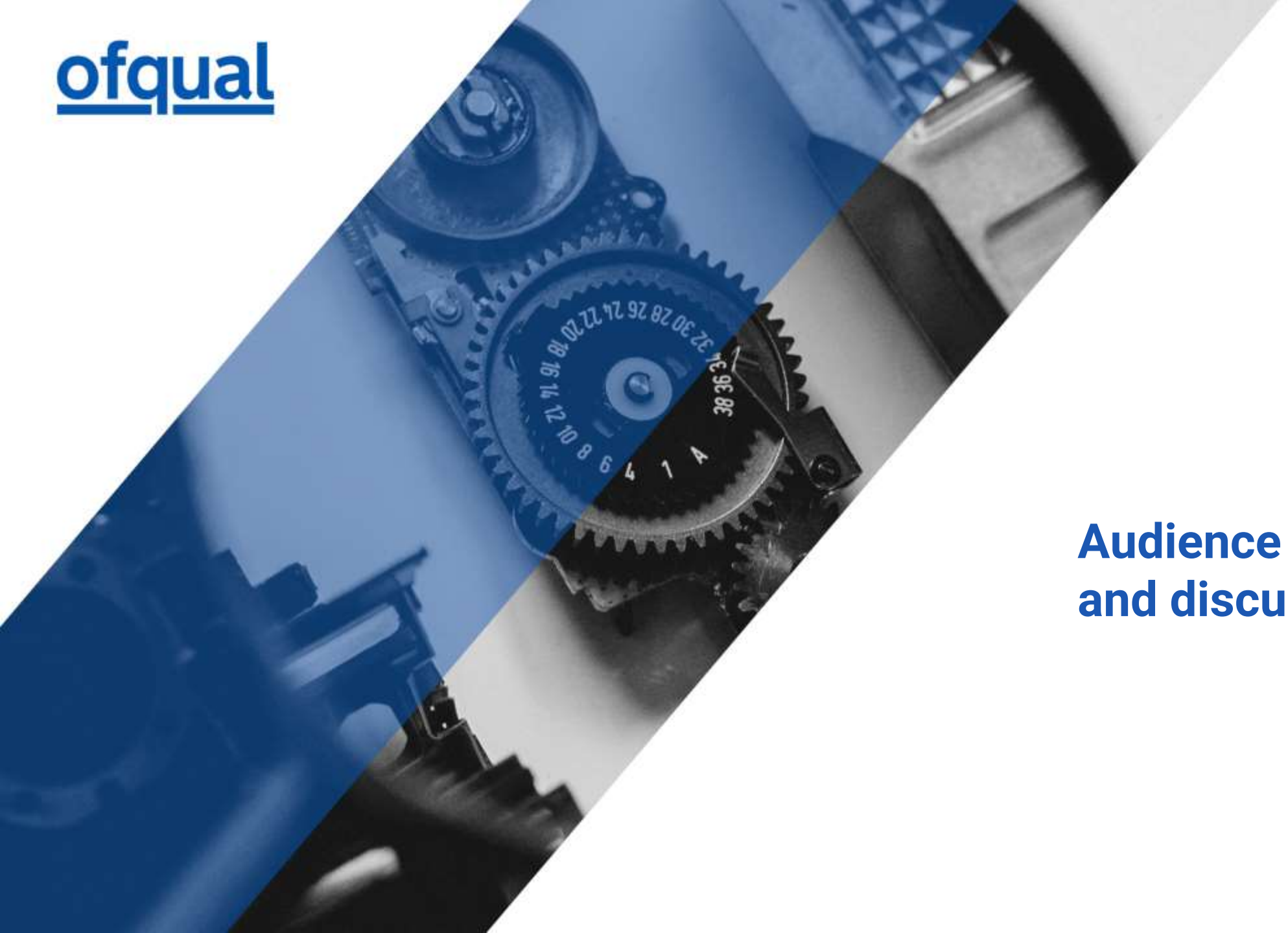
Abstract

To help achieve national consistency of assessment and reporting in the Australian Vocational Education and Training sector, there is a need to develop a set of national principles for graded performance assessment. This paper challenges a number of prevailing principles from both a theoretical and assessment perspective, namely that grades must be criterion referenced (Williams & Bateman, 2003), meaningful (Rumsey, 2003) and applied once competence has been achieved (Williams & Bateman, 2003). This paper argues that the use of generic criteria cannot be defended in terms of their validity and reliability and that a clear understanding of the underlying developmental continuum of learning is required to inform the development of meaningful and valid criteria and descriptors of quality performance. Finally, the paper proposes a set of principles that have been grounded in theory, have been put to the test in large-scale research, and are consistent with international literature on competence assessment.

Introduction

Since the introduction of competency based training and assessment into the Australian Vocational Education and Training (VET) sector in the early 1990s, the way in which the outcomes of assessment should be reported has been a contentious issue amongst researchers, policy makers and practitioners. In the early 1990s, debate typically focused on whether the principles that underpinned competency based assessment implied only one acceptable standard of performance or whether

Should we necessarily assume that there are important, intrinsic differences between grading in technical and vocational education and training contexts versus grading in other contexts (general education)?



Audience reactions and discussion

ofqual

Experiences of vocational and technical grading - practitioner views

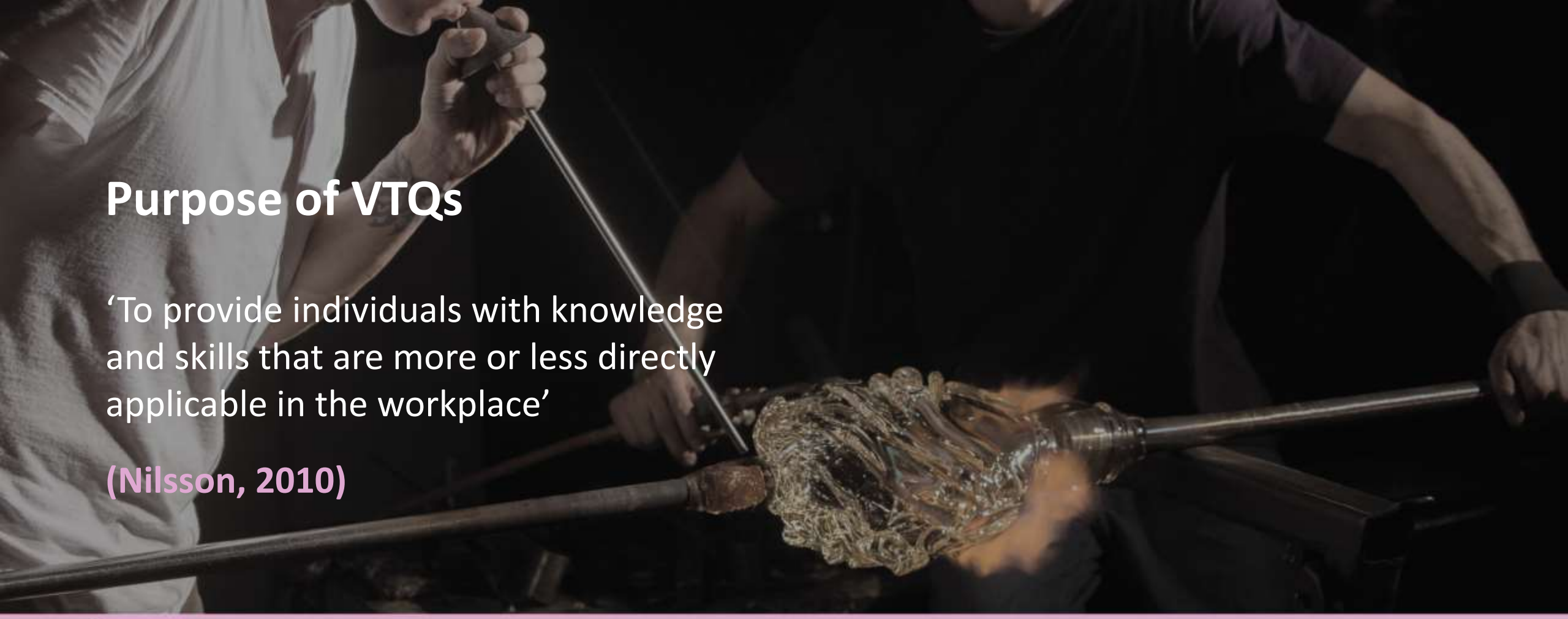
AO perspectives on grading design for the assessment of practical skills.

Carina Fagan

Grading Design in VTQs for the Assessment of Practical Skills

An awarding organisation's perspective

Carina Fagan
Chief Academic Officer, VTCT



Purpose of VTQs

‘To provide individuals with knowledge and skills that are more or less directly applicable in the workplace’

(Nilsson, 2010)

VTQs should also focus on connections between highly specific technical content and content with broader applicability.

Clinical Skills Grading

Objective Structured Clinical Examination (OSCE) – How discriminatory is an OSCE?

OSCE Global Score Descriptors – Queen's University, Belfast

Excellent

Excellent performance of skill. Outstanding demonstration of technical and non-technical aspects of skill.
Air of confidence and fluent.

Clear pass

Acceptable performance of skill. Despite omissions/errors demonstrated in performance of skill – safe to progress.
At times can be formulaic in approach.

Clear fail

Performance of skill did not come up to a passing standard. Appeared disorganised. Unsafe and unsuitable to progress.

Very good pass

Very good performance of skill. Majority of the technical aspects of the skill demonstrated. Few minor and non-essential omissions/errors. Examiner more than satisfied that candidate has passed station.

Borderline: 'pass doubtful'

Patchy performance of skill. Examiner undecided whether to pass or fail candidate. Demonstrated some aspects of the skill however omissions and inaccuracies occurred in their performance of the skill. Often formulaic in approach and struggled with performing skill.

VTCT

Purpose of VTQs

Consultation with stakeholders:

Employers

Competition organisers (World Skills)

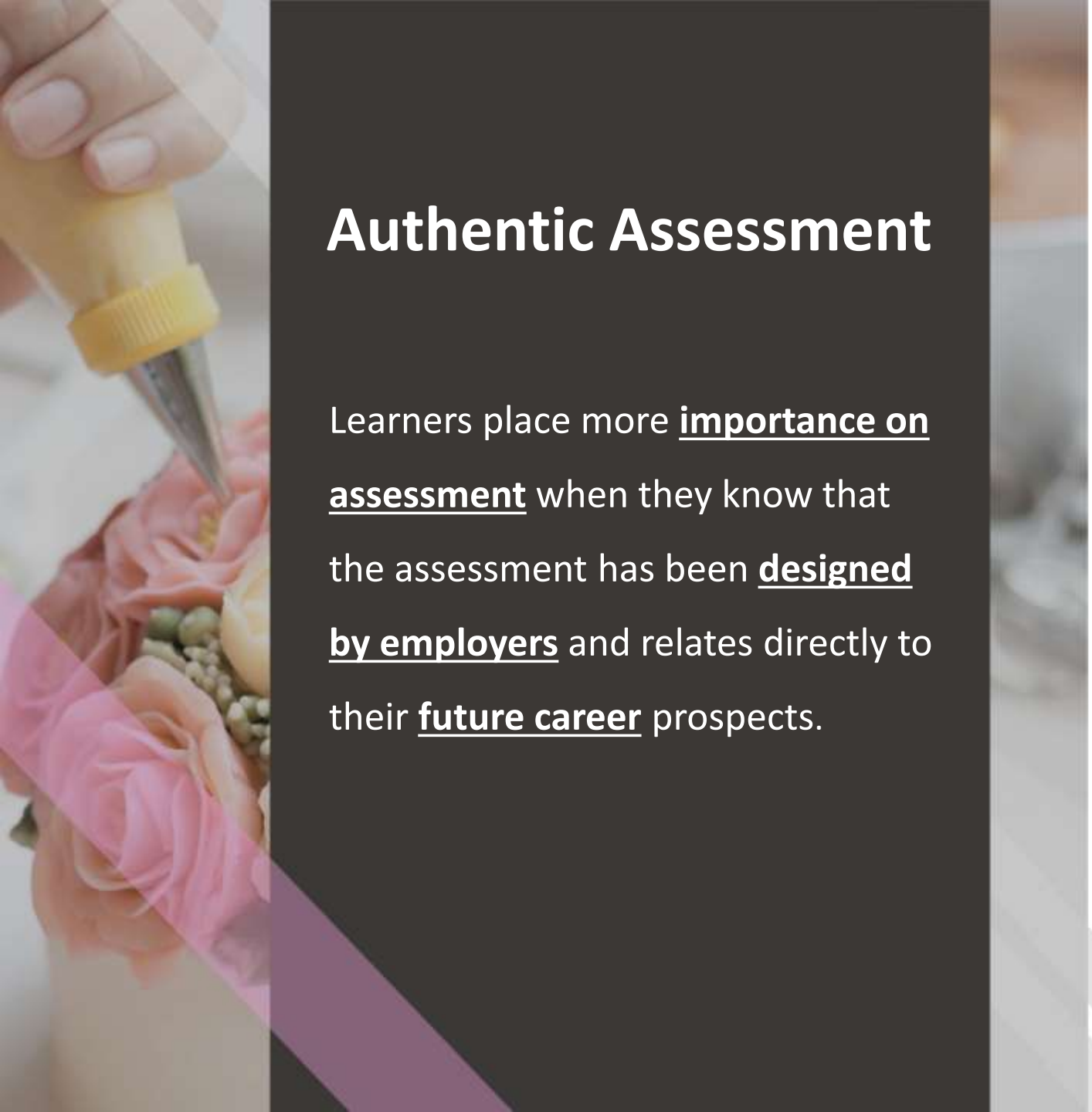
Learners

Validity

- The degree to which it is possible to **measure**
- Whatever **needs** to be **measured**
- **In accordance with the purpose of the qualification**

Grading of VTQs

- **Measure** whatever **needs** to be **measured**
- **Grade** whatever **needs** to be **graded**
- **In accordance with the purpose of the qualification**



Authentic Assessment

Learners place more **importance on assessment** when they know that the assessment has been **designed by employers** and relates directly to their **future career** prospects.

‘Tasks are replicas of the kinds of problems faced by professionals in the field.’

(Wiggins, 1993)

‘Connections to real-life skills.’

(Meyer, 1993)

‘When learners perceive that assessment tasks resemble their future professional practice, they are more motivated and invested in the development of their knowledge and skills.’

(Guilkers, 2007)

VTCT

Stakeholder Values

Employers told VTCT that any VTQ grading model design must be underpinned by a strong stakeholder-determined methodology:

- ✓ Grades must be **representative** of the skills and knowledge that are of relative importance to employers and the purpose of the target job role.
- ✓ Grades must **differentiate** proportionately between occupation-specific technical skills, applied knowledge and employability attributes.
- ✓ Grades must provide **transparent** information at the component level about prospective employees technical strengths and weaknesses, as well as an overall qualification grade.

VTCT

Assessment Methodology

Analytical and holistic assessments have strengths and weaknesses:

- Reliability
- Logistical complexity
- Compensation
- Coverage
- Face validity
- Learner engagement
- Cost

A mixed method approach using both analytical and holistic assessments within the one qualification could help to ensure that the weaknesses of any one assessment method is minimised.

Component weighting representative of value placed by employers on the topic area.

Analytical Grading Rubrics

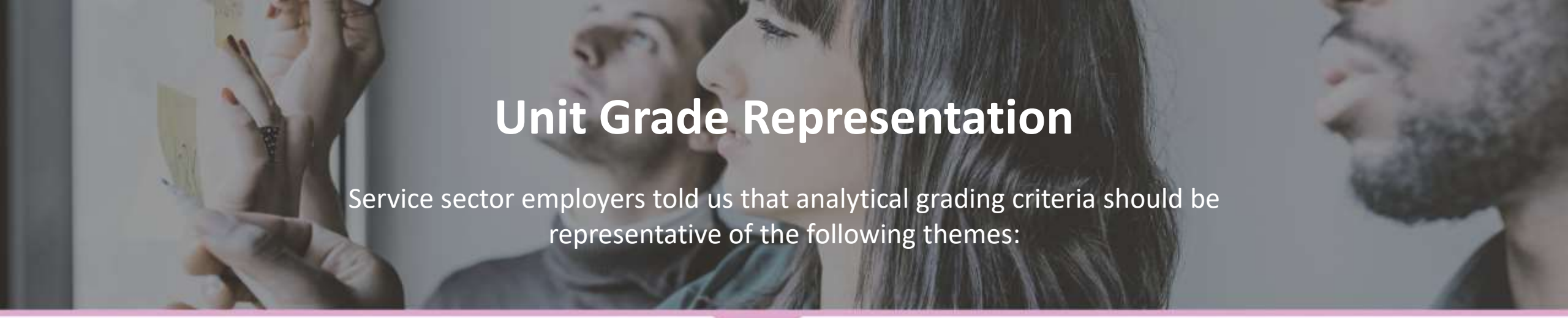
Grading at unit level:

- Detailed and analytical interrogation of core technical skills and applied knowledge; proficiency and mastery
- Ephemeral evidence – criterion-referenced
- Non-compensatory and hurdle based
- Measure interaction, creativity and adaptability
- Measure higher-order cognitive skills such as justification, evaluation, reflection

Mandatory units – Skills and knowledge of primary and core (occupation-specific) importance to employers, should contribute to the overall grade.

Optional units – Skills and knowledge of secondary or niche importance to employers should not contribute to the overall grade.

VTCT



Unit Grade Representation

Service sector employers told us that analytical grading criteria should be representative of the following themes:

Pass criteria

Proficient technical skill and safe performance.

Merit criteria

Excellent use of techniques and technical processes, plus a commercial focus e.g. time management, organisation skills, promoting and selling additional products and services.

Distinction criteria

Experienced-worker standard (EWS) of proficiency demonstrated by an excellent finished result (product evidence), plus justification of methods selected, critical evaluation and reflection.

VTCT

Holistic Grading Rubrics

Grading a synoptic performance using holistic assessment methods

- Testing the integrated application of skills and knowledge
- Enhancing the links between units
- Combining different elements of learning
- Focussing on broader employability
- Best fit global judgement (no hurdles)

Contextualised global judgement:

- Context (salon, kitchen, gym)
- Process (range of services and techniques)
- Outcomes (satisfied clients and customers)

VTCT

Synoptic Grade Representation

Service sector employers told us that holistic/global synoptic grading criteria should be representative of the following themes, whereby the assessor makes an impressionistic or intuitive (Sadler, 2008) global judgement, which is by nature compensatory.

Pass global judgement descriptor - Proficient technical skill and safe performance.

Merit global judgement descriptor - Customer satisfaction plus commercial skills (contributing to the development of the business).

Distinction global judgement descriptor - All of the above plus demonstration of personal reflection and professional development.

VTCT

Grading Framework and Design Principles for Assessing Practical Skills and Applied Knowledge in VTQs

UNIT GRADED ANALYTICAL PRACTICAL ASSESSMENTS – 50%

Mandatory units graded P/M/D with each having a points value.

Optional units graded P/F with no points value.

+

EXTERNAL THEORY EXAM – 30% of overall qualification grade

+

SYNOPTIC GRADED HOLISTIC PRACTICAL ASSESSMENT – 20%

Mandatory units only included in the synoptic holistic practical assessment P/M/D with a points value.

VTCT

Awarding Organisation Reflections

- Grade what employers want to be graded – purpose of the qualification
- Use a mixed methods approach – combine analytical assessment of isolated technical skills and applied knowledge with holistic global judgements on the whole performance to test broader skills and employability attributes
- Aggregation and weighting based on relative-importance determined by stakeholders
- Assessments should properly balance the impact of the work (product evidence) with assessment of process (methods and techniques)
- Grading in VTQs should be **representative** ('the grade should tell a story') rather than quantitative
- Further research and nationally agreed design principles for grading VTQs

VTCT

ofqual

Observations on apprenticeships and grading

Barry Smith

The National Skills Academy
RAIL

Observations on apprenticeships and grading

Note of caution

- This is about the apprenticeship reforms
- It is CBA but not as we knew it.
- We are in an area where we often use terminology without precision:
  - TVET
  - VTQs
  - VQs
  - NVQs
  - CBA

EQA, apprenticeships and the story so far

The National Skills Academy RAIL

What's changed?

- Continuous, unit-based assessment for end-point assessment
- Local/familiar, flexible and low-threat assessment
- Distant/unfamiliar, fixed and high-stakes assessment
- End-point assessment market place with new players
- System framed by (national) occupational standards is now framed by occupational job roles
- Introduction of Grading into CBA
- Employers activated as a producer and consumer of product as demand-side is empowered/prioritised.

EQA, apprenticeships and the story so far

Grading is contested in this space

- Battle of the binary – competent or not - ‘end of …!’
- Push to embrace and innovate – getting back to first principles about what is valued
- Pressure to ensure a measure of consistency but do we sometimes risk confusing consistency in assessment with conformity in assessment

EQA, apprenticeships and the story so far

The National Skills Academy RAIL

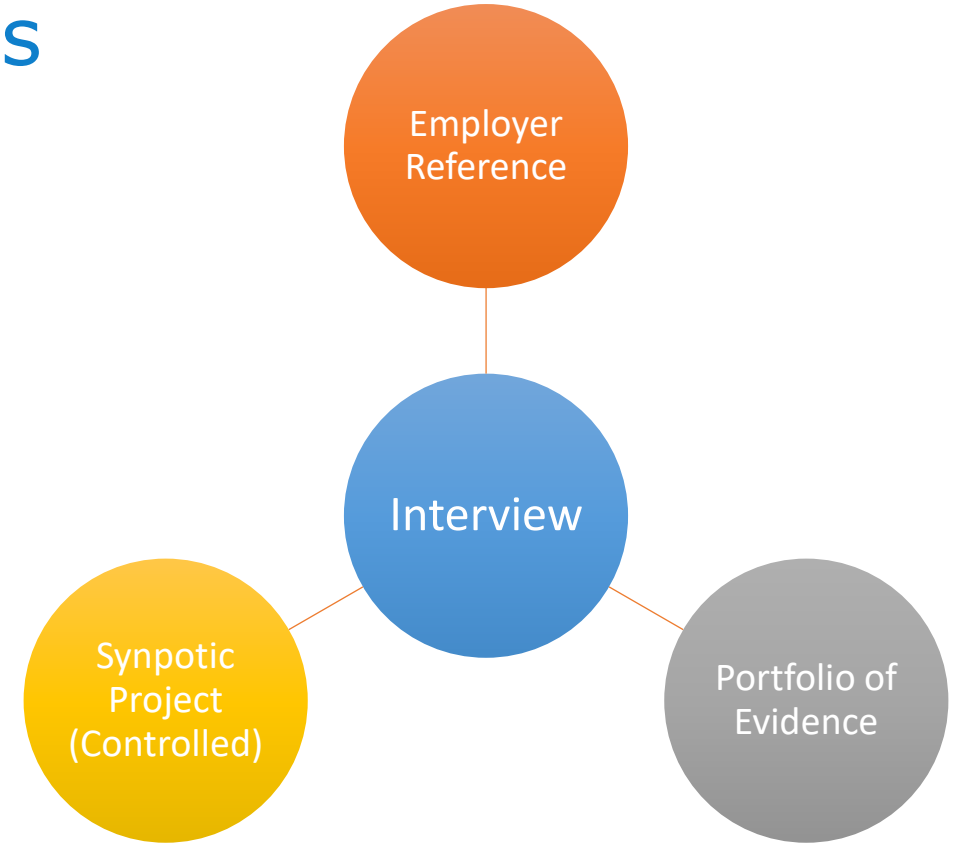
Grading - The inconvenient truth

It (grading) happens routinely but informally in many competence-based assessment environments.

Example of Innovation

The National Skills Academy RAIL

Digital Industries



Example of Innovation

The National Skills Academy RAIL

Digital Industries

Pass	Merit		Distinction
The What, the How and the With Whom are all at the required level	The What and the How are each significantly above the required level	The What and the With Whom are each significantly above the required level	The What, the How and the With Whom are significantly above the expected level

A bit more challenging

The National Skills Academy RAIL

Knowledge, Skills and Behaviour assessed	Description	Assessment method (Knowledge Test, Project Presentation or Portfolio Interview)	Fail: below <60%	Pass: between 60-79%	Distinction: above 79%>
Record and document production	Produces accurate records and documents including: emails, letters, files, payments, reports and proposals. Makes recommendations for improvements and present solutions to management. Drafts correspondence, writes reports and able to review others' work. Maintains records and files, handles confidential information in compliance with the organisation's procedures. Coaches others in the processes required to complete these tasks.	Portfolio interview	• Frequent mistakes in written documentation, requiring regular correction • Zero or very few workable recommendations are made • Communications not kept confidential	• Records are accurate, rarely require correction and are treated confidentially • Recommendations and solutions only need minor improvements • Supports others in producing documents and can provide examples	• Records are consistently accurate and confidential • Recommendations are insightful, clearly recorded and results in a clear benefit to the organisation • Offers to coach others and good performance is recorded in feedback

EQA, apprenticeships and the story so far

The National Skills Academy RAIL

Grading - 'Could do better'

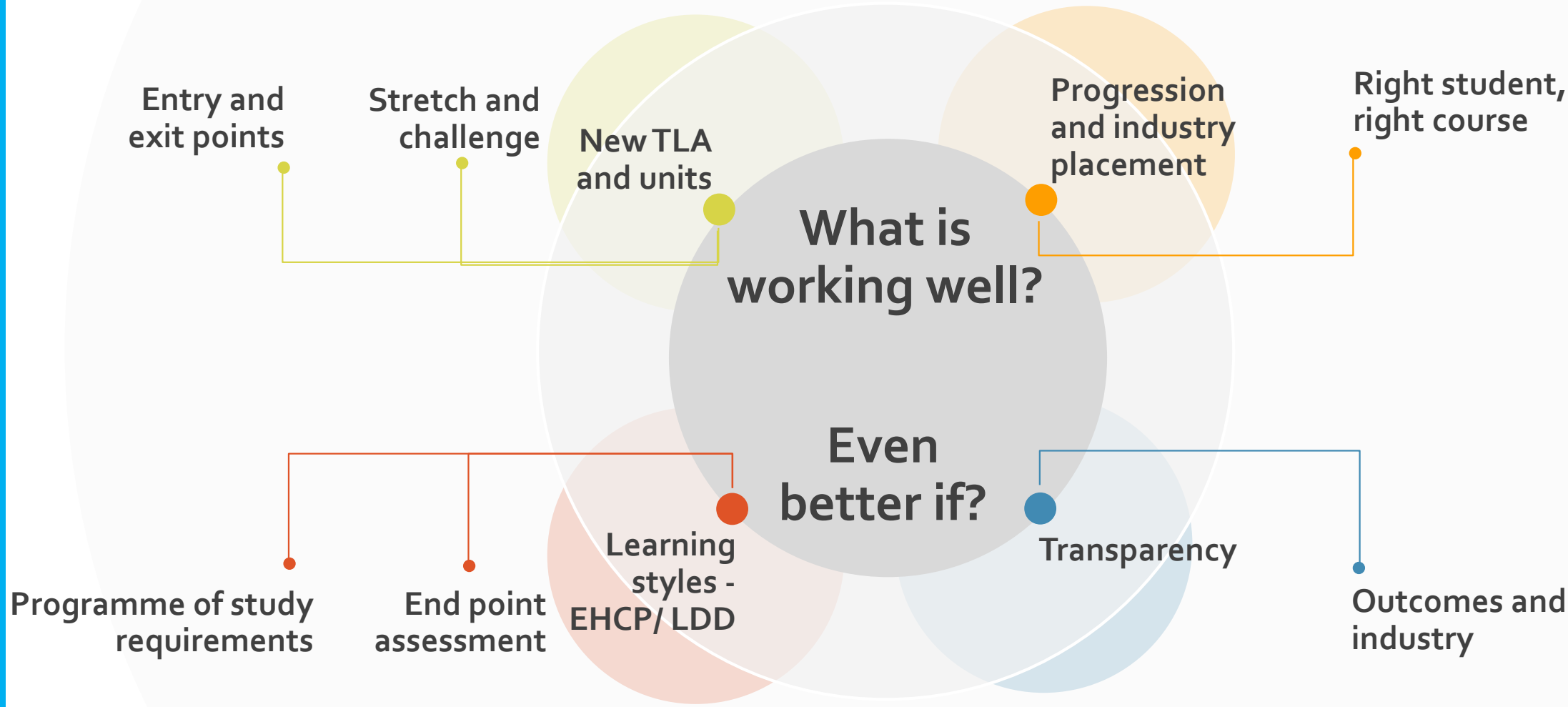
- More focus on:
  - communicating the worth of grading, but
  - find the intrinsic worth and drivers for grading in the occupation
- Resist the tendency to get caught up in the gravitational pull exerted by general grading approaches
- Add the right sort of value where focus is on making what's new work and not on changing it to be familiar and to conform

ofqual

Grading within a Further Education College

Adam Sturt

Grading within a Further Education College



Practitioner experiences of grading

- Carina Fagan, Chief Academic Officer, VTCT
- Barry Smith, National Skills Academy for Rail
- Adam Sturt, City College Norwich

- Chair – Phil Beach for audience discussion and reactions

ofqual

Lunch
12.30-1.30pm

ofqual

Grading of Vocational and Technical Assessments

Mary Ward House
8d0da34dfe

slido.com
#ofqual

ofqual

Reflections of grading vocational assessments

The use of frameworks for grading: let a hundred frameworks Bloom!

Professor Steve Higgins

The use of frameworks for grading: let a hundred frameworks Bloom!

Professor Steven Higgins
School of Education
Durham University
s.e.higgins@durham.ac.uk
@profstig

Durham University

Let a hundred frameworks Bloom?

1. Dominance of Bloom
2. Overview of ‘Frameworks for thinking’
3. Two examples

Dominance of Bloom

Taxonomy of Educational Objectives:

Vol 1 The Cognitive Domain

Bloom, **Englehart, Furst, Hill, & Krathwohl** (1956)

Behaviourist psychology

Sees thinking skills as generalizable

No systematic rationale

Not empirically based or validated

What about the other domains?

1. The cognitive domain (knowledge-based)
- 1.1 Knowledge
- 1.2 Comprehension
- 1.3 Application
- 1.4 Analysis
- 1.5 Synthesis
- 1.6 Evaluation

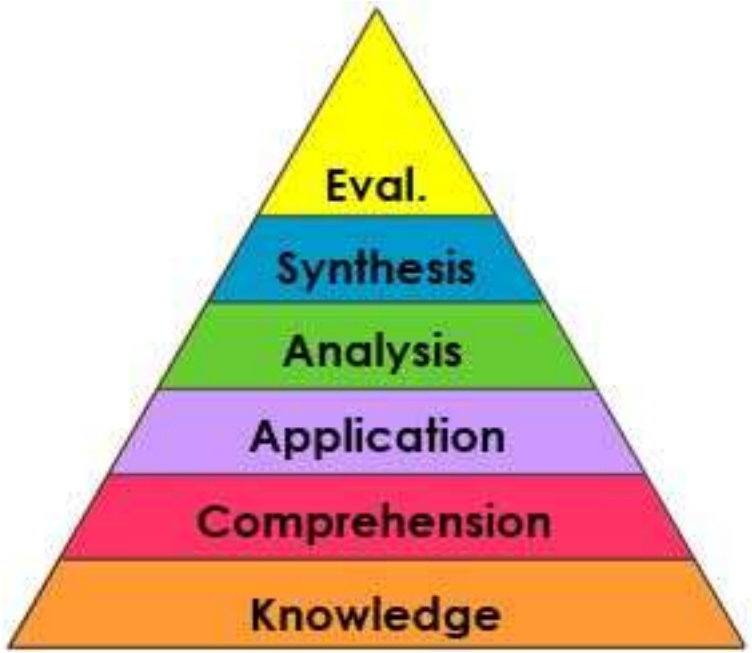
2 The affective domain (emotion-based)
- 2.1 Receiving
- 2.2 Responding
- 2.3 Valuing
- 2.4 Organizing
- 2.5 Characterizing

3 **The psychomotor domain (action-based)**
- 3.1 Perception
- 3.2 Set
- 3.3 Guided response
- 3.4 Mechanism
- 3.5 Complex overt response
- 3.6 Adaptation
- 3.7 Origination

Appendix 1: additional knowledge taxonomy

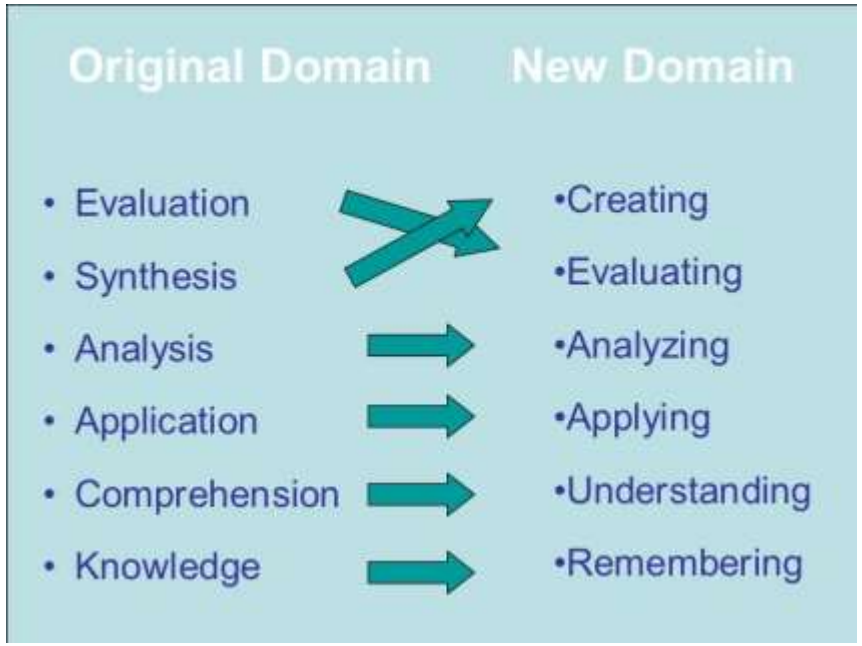
Knowledge
- 1.10 Knowledge of specifics
  - 1.11 Knowledge of terminology
  - 1.12 Knowledge of specific facts
- 1.20 Knowledge of ways and means of dealing with specifics
  - 1.21 Knowledge of conventions
  - 1.22 Knowledge of trends and sequences
  - 1.23 Knowledge of classifications and categories
  - 1.24 Knowledge of criteria
  - 1.25 Knowledge of methodology
- 1.30 Knowledge of the universals and abstractions in a field
  - 1.31 Knowledge of principles and generalizations
  - 1.32 Knowledge of theories and structures

Important revision



Bloom et al. (1956)

Anderson et al. (2001)



1. Bloom

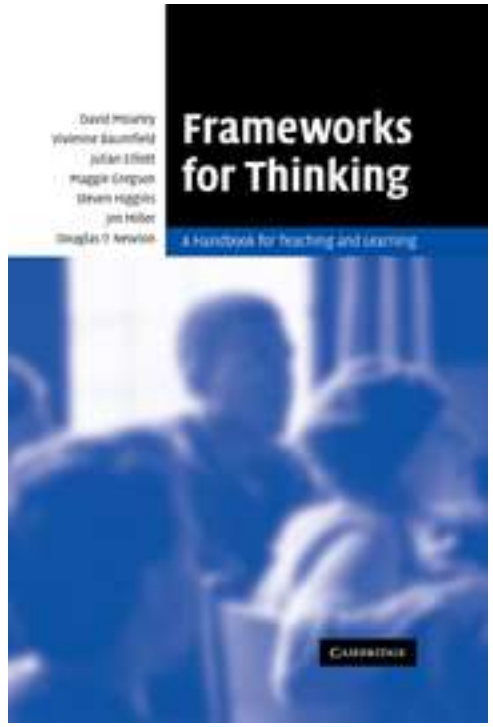
Frameworks for thinking and learning

Report to LSDA (2004)

- identified over 60 frameworks
- evaluated 35 in-depth
- Published by CUP as ‘Frameworks for thinking’ (Moseley *et al.* 2005)

‘Families’

- Instructional design frameworks
- Critical and ‘productive’ thinking frameworks
- Explanatory models of cognitive structure and/or development
- All-embracing frameworks (covering personality, thought and learning)



Instructional design frameworks

Bloom's taxonomy of educational objectives (cognitive domain)
Feuerstein's theory of Mediated Learning through Instrumental Enrichment
Ausubel and Robinson's six hierarchically ordered categories
Gagne´'s eight types of learning and five types of learned capability
Hannah and Michaelis's comprehensive framework for instructional objectives
Williams' model for developing thinking and feeling processes
Biggs and Collis' SOLO taxonomy
Quellmalz's framework of thinking skills
Stahl and Murphy's domain of cognition taxonomic system
Presseisen's models of basic, complex and metacognitive thinking skills
Anderson and Krathwohl's revision of Bloom's taxonomy
Gouge and Yates' ARTS Project taxonomies of Arts Reasoning and Thinking Skills

Critical and ‘productive’ thinking frameworks

Altshuller’s TRIZ Theory of Inventive Problem Solving

Lipman’s three modes of thinking and four main varieties of cognitive skill

Baron’s model of the good thinker

Ennis’ taxonomy of critical thinking dispositions and abilities

Gubbins’ taxonomy

Halpern’s reviews of critical thinking skills and dispositions

Paul’s model of critical thinking

Jewell’s reasoning taxonomy for gifted children

Explanatory models of cognitive structure and/or development

Piaget's stage model of cognitive development
Guilford's structure of intellect model
Perry's developmental scheme
Belenky's '*Women's Ways of Knowing*' developmental model
Gardner's theory of Multiple Intelligences
King and Kitchener's reflective judgment model
Koplowitz's stages in adult cognitive development
Carroll's three-stratum theory of cognitive abilities
Pintrich's general framework for self-regulated learning

All-embracing frameworks (covering personality, thought and learning)

Romizowski's analysis of knowledge and skills

Hauenstein's conceptual framework for educational objectives

Wallace and Adams' Thinking Actively in a Social Context

Jonassen and Tessmer's taxonomy of learning outcomes

Vermunt and Verloop's categorisation of learning activities

Marzano's new taxonomy of educational objectives

Sternberg's model of abilities as developing expertise

Romiszowski's analysis of knowledge and skills

Table 6.1. Romiszowski's knowledge categories

1.1 *concrete facts*
- 1.1.1 concrete associations (things observed and remembered)
- 1.1.2 verbal (symbolic) information (including all knowledge of a factual nature that has been gained by means of a symbolic language)
- 1.1.3 fact systems (structures or schemata)

1.2 *procedures*
- 1.2.1 linear procedures (chains)
- 1.2.2 multiple discriminations (distinguishing similar information)
- 1.2.3 algorithms (procedures which may be complex but which guarantee successful performance if followed correctly)

2.1 *concepts*
- 2.1.1 concrete concepts (classes of real objects or situations)
- 2.1.2 defined concepts (concepts which are classes of other concepts and cannot be learned without the use of a suitable language)
- 2.1.3 concept systems (structures or schemata)

2.2 *principles*
- 2.2.1 rules of nature (principles we can observe to be in operation in the world either by direct observation of by inference from their effects)
- 2.2.2 rules of action (general heuristics regarding the appropriate actions or reactions to specific situations)
- 2.2.3 rule systems (theories or strategies suitable for a given class of problems).

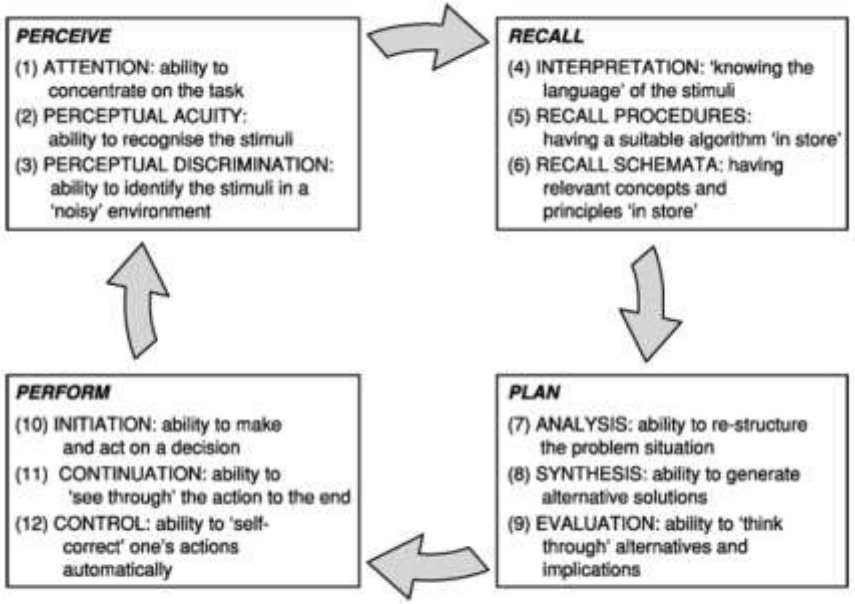


Fig. 6.1. Romiszowski's skill cycle.

3 . Examples

Table 3.4. The SOLO taxonomy levels, with descriptors and criteria

SOLO description	Capacity	Relating operation	Consistency and closure
Pre-structural	Minimal: cue and response confused	Denial, tautology, transduction. Bound to specifics.	No need felt for consistency: closure without seeing the problem.
Unistructural	Low: cue and one relevant datum	Can generalise only in terms of one aspect.	No need felt for consistency: thus closed too quickly; jumps to conclusions so can be very inconsistent.
Multistructural	Medium: cue and isolated relevant data	Can generalise only in terms of a few limited and independent aspects.	Feeling for consistency: closure too soon on basis of isolated fixations so can reach different conclusions with same data.
Relational	High: cue and relevant data and interrelations	Induction: can generalise within given or experienced context using related aspects.	No inconsistency in given system, but closure is unique to given system.
Extended abstract	Maximal: cue and relevant data and interrelations and hypotheses	Deduction and induction: can generalise to situations not experienced.	Inconsistencies resolved: no need for closed decisions; conclusions held open or qualified to allow logically possible alternatives.

SOLO Taxonomy: relational complexity

3 . Examples

References

Bloom, B., Englehart, M, Furst, E., Hill, W. & Krahtwohl, D. (1956) *Taxonomy of Educational Objectives: The Classification of Educational Goals. Handbook 1: Cognitive Domain*. New York: Longmans Green, 1956.

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, P. R., Raths, J., et al. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives* NewYork: Allyn & Bacon.

Moseley, D., Baumfield, V., Elliott, J., Higgins, S., Miller, J. & Newton D. P. (2005) *Frameworks for thinking: a handbook for teaching and learning* Cambridge: Cambridge University Press ISBN 0521848318.

ofqual

Getting clearer about the use of criteria

Professor Steve Higgins

Getting clearer about the use of criteria

Dr Gerard Lum
King's College London

Improving on Competence/Outcomes-Based Assessment

Questionable assumptions relating to:

- the (binary) nature of occupational capability/ expertise
- facility of language to ‘capture’ human capabilities in precise terms
- the nature of the judgements involved in the process of assessment (objective/correspondence)

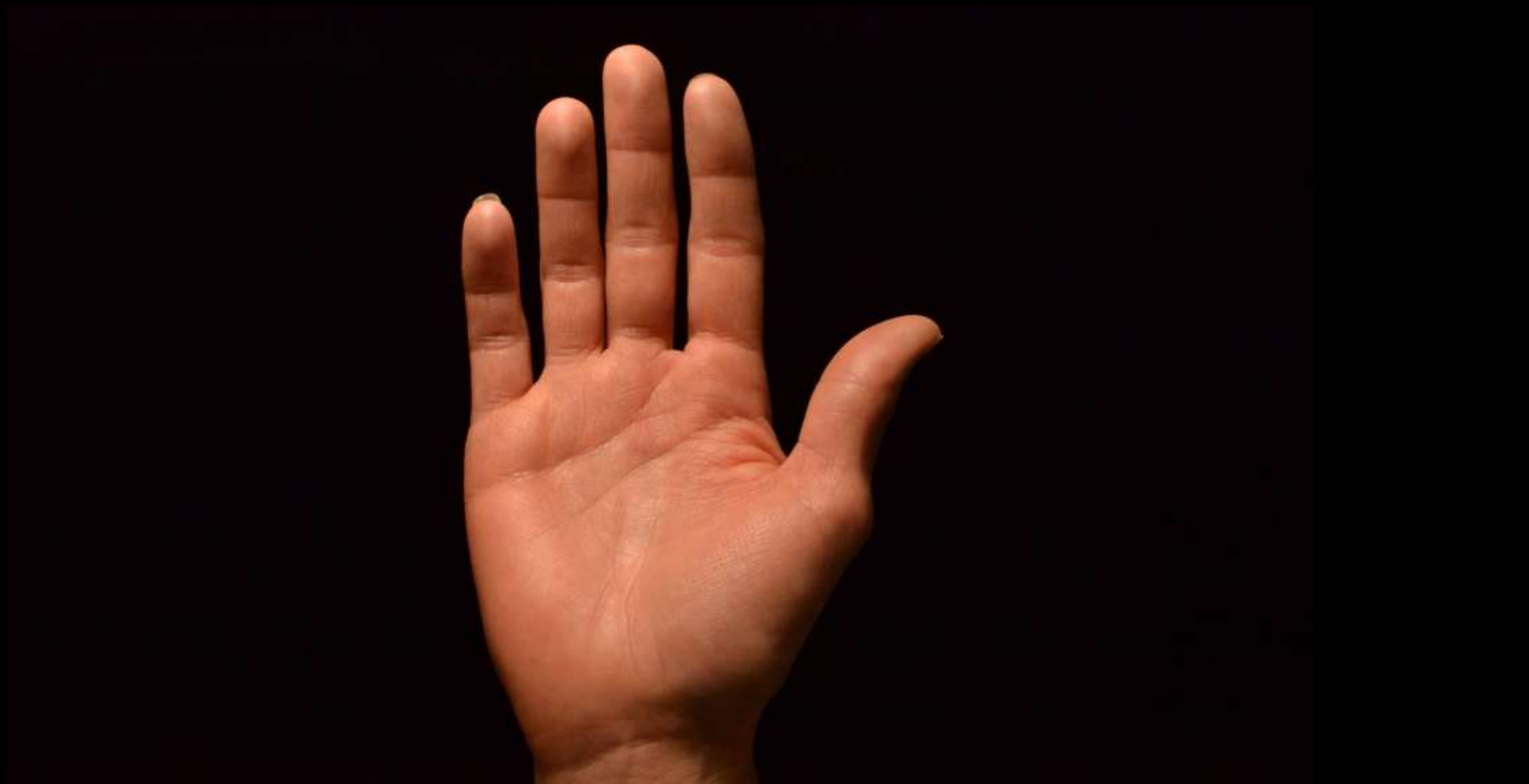
‘Precision in the use of language’

Statements must **accurately** communicate their intent. For accurate communication of the outcomes of competence and attainment, a **precision** in the use of language in such statements will need to be established, approaching that of a **science**. The overall model stands of falls on how effectively we can state competence and attainment.

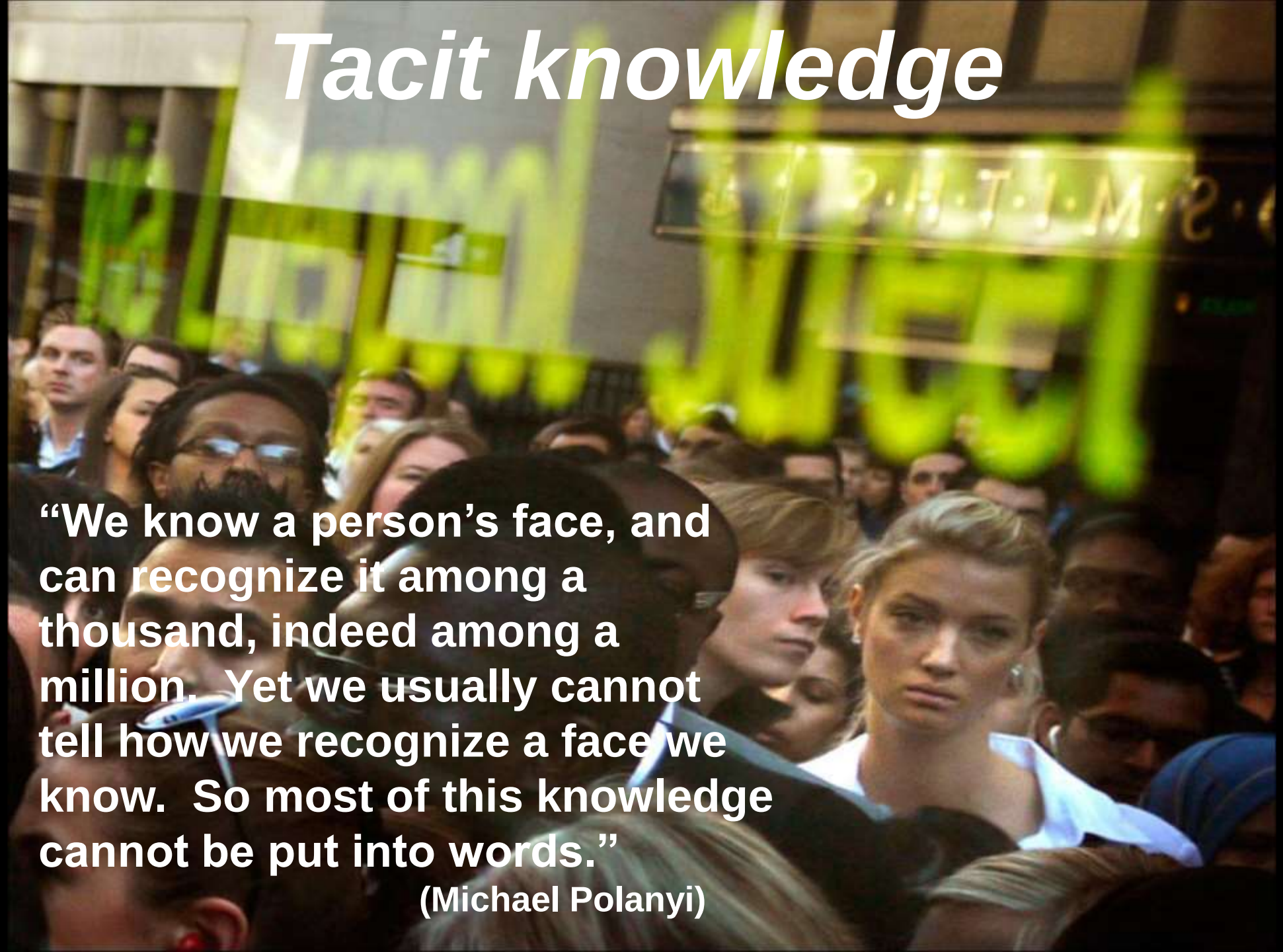
(Jessup, 1991)

Are *complete* descriptions possible?

(Waismann, 1951)



Tacit knowledge



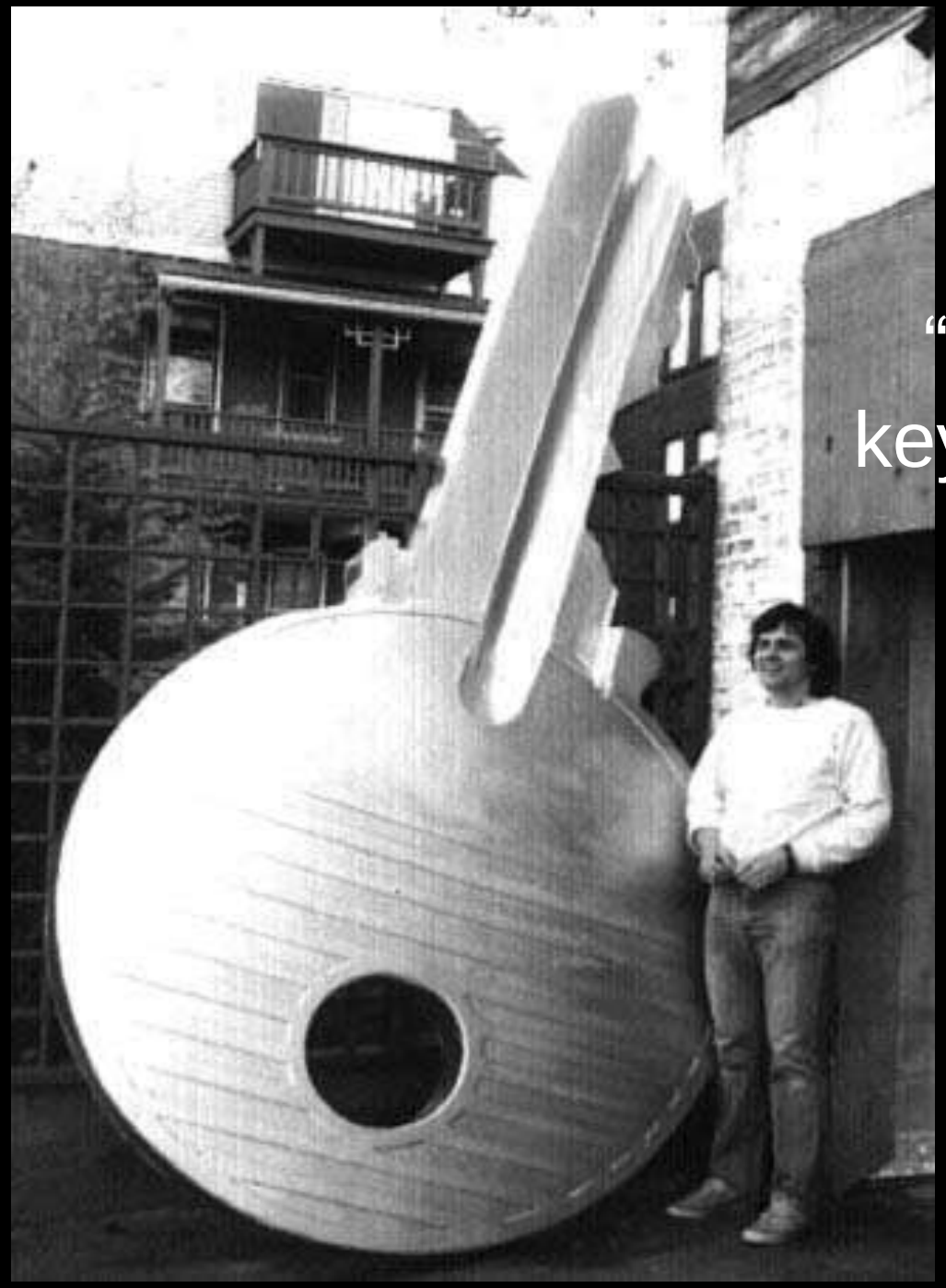
“We know a person’s face, and can recognize it among a thousand, indeed among a million. Yet we usually cannot tell how we recognize a face we know. So most of this knowledge cannot be put into words.”

(Michael Polanyi)

‘...imagine a group of highly trained observers describing an American football game in statements only of brute facts. What could they say by way of description?’ (John Searle)

“She handed him the key
and he opened the door”

“She handed him the key and he opened the door”



“She handed him the key and he opened the door”

“She handed him the key and he opened the door”

...what does he know?

Behaviour as against mind?

The Right/Wrong Scenario

Same questions given...

Same responses sought...

Same responses obtained...

... different judgements made

The Right/Wrong Scenario

The choice:

a) *specified* behaviour/evidence

or

b) all available evidence

Difference consists in the stance taken towards ...

Two kinds of assessment

1. ***Prescriptive*** mode:

- strictly prescribed evidence
- binary results
- low stakes use only

Two kinds of assessment

2. ***Expansive*** *mode:*

- considers *all* relevant evidence
- results a matter of degree
- must be used when stakes are high

The necessity to make a clear choice...

Prescriptive mode

Expansive mode

... for validity and reliability

Ontologically differentiated criteria and mode shift

Assessment focus → **(Child's ability to) write imaginative, interesting and thoughtful texts**

Band Criteria →

- a chronological sequence of events
- time-related words or phrases
- some recognisable letters
- a simple series of occurrences
- a list of changes
- a description of the pebble

The Legal Analogy

... actus reus and mens rea

How do law courts treat evidence?

Prescriptive or Expansive Mode?

Innocent or guilty?

Fit to practise?

Only <u>expansive</u> mode assessment will suffice

Some broad conclusions :

1. We cannot assume that criteria /descriptions sufficiently determine the capabilities at issue - we need to think about the understandings implicitly required for their interpretation.
2. A clear distinction should be made between judgements of identity and judgements of significance, i.e. between prescriptive and expansive modes of assessment.
3. To the extent that assessment has high stakes then there will need to be a place for expansive mode assessment.
4. There is an important role for the graded assessment of ‘inputs’ in aggregated procedures.

Grading in medical education: managing the tensions in a competency-based discipline

Dr Matthew Homer

School of Medicine
FACULTY OF MEDICINE AND HEALTH

UNIVERSITY OF LEEDS

Grading in medical education:

Managing the tensions in a competency-based discipline

Matt Homer

Leeds Institute of Medical Education (LIME)

University of Leeds

@LeedsARG

- **Overview of talk**
- *Grading in medicine in the UK*
- Brief background
- Undergraduate – in-course, summative, workplace, degree certification – and wider assessment trends
- National policy - forthcoming Medical Licensing Assessment (MLA)
- Conclusions

- **Background**
- Medical education is high-stakes
- Need to make the right decisions in terms of progression, and access to the profession
- Wide range of stakeholders – students, patients, the community, the health care professions, the tax-payer…
- Competency-based and criterion referenced assessment
- Psychometrics and research – assessment/standard setting
- Valuing professional judgment – embracing the subjective

- **Undergraduate degree**
- MBChB (*Bachelor of Medicine and Bachelor of Surgery degrees*) – 5 years at Leeds
- Typically:
- In-course assessments – pass/fail – Assessment for Learning
- Applied Knowledge Tests from early years – standard set – Angoff/Ebel/Rasch(?) and graded (norm referenced?)
- Clinical performance examinations in later years – OSCEs (Objective Structured Clinical Examinations) – more later.
- Workplace-based assessments – more later

- **Scoring in OSCEs**
- Students rotate around a set of clinical encounters (‘stations’: n ≈ 13-26)
- A ‘patient’ and a trained assessor (clinician) – both judge the performance
- Patient rating – how likely would you want to see this doctor again? – mostly for formative purposes
- Assessor – a holistic judgment (‘global grade’), and a checklist score
- A set of generic global grade descriptors – ‘fail’, ‘borderline’, ‘clear pass’, ‘good pass’ and ‘excellent pass’

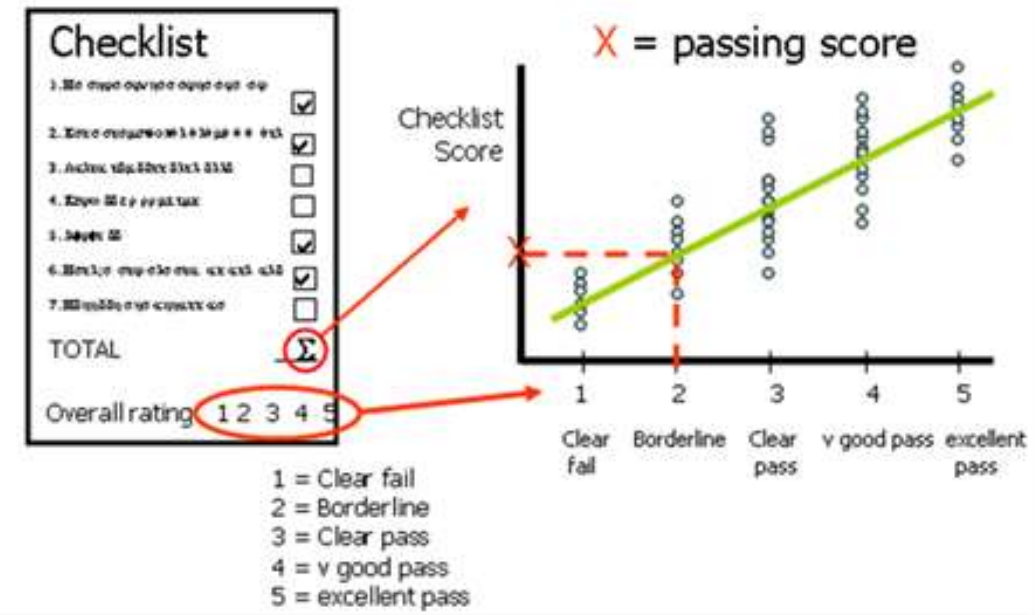
Grade descriptors - generic but detailed – select examples

- ***Clear Fail***
- Little idea of how to…
- Disorganized….
- Unable to….
- ***Clear Pass***
- Systematic overall approach…
- Demonstrates sufficient…
- Able to…
- ***Excellent***
- Overall superior approach…
- Flexible, adaptive…
- High levels of…

- Important for assessors to have a shared understanding of key levels of performance
- Minimise unwanted assessor variation e.g. ‘hawks and doves’

Passing/failing in individual stations

- All checklist scores and grades are combined to create a station cut-score
- Borderline regression method (Pell 2010)
- Lots of ‘metrics’ to investigate and monitor ‘quality’
- Station-level pass/fail decision based on **checklist score alone**

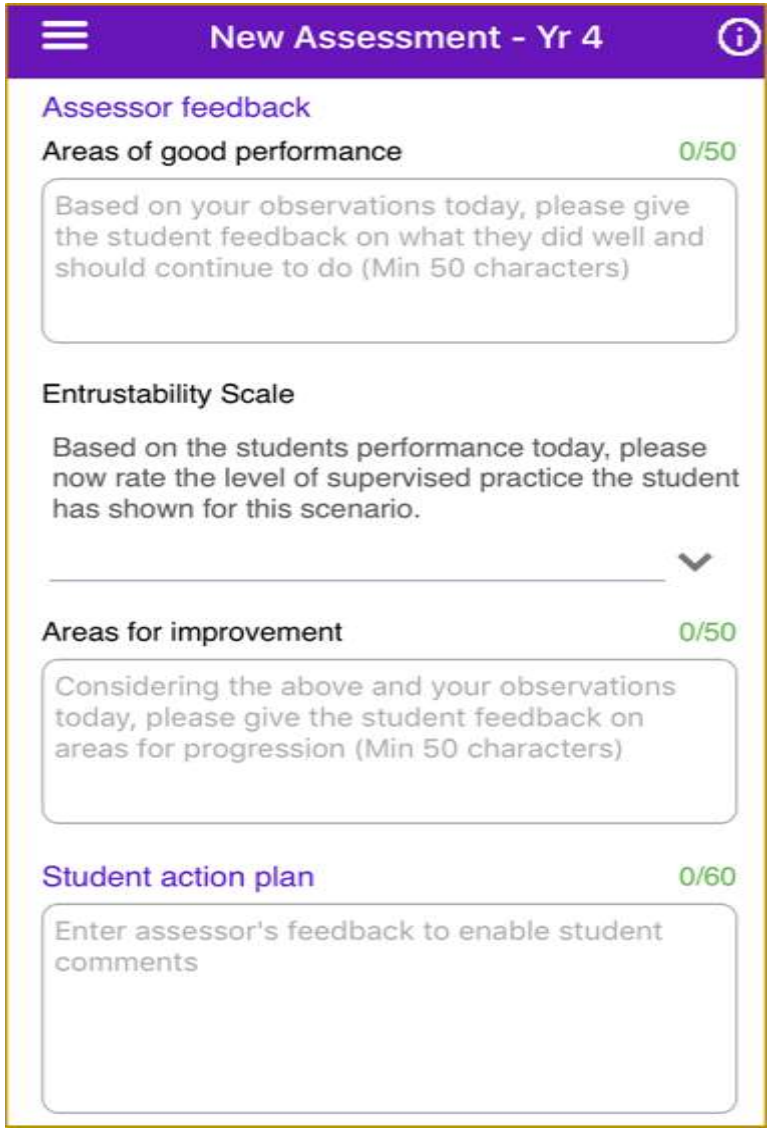


- **Overall OSCE pass/fail decision and grade**

- Individual station cut-scores aggregated
- Subjectivity – ‘hawks and doves’ – balances out?
- This is the overall cut-score for the exam (plus a standard error of measurement – Hays 2008)
- Also, conjunctive standard – minimum no. of stations passed – to limit excessive compensation across stations
- Final grade (A ‘excellent’,…, to E ‘fail’) – based on performance relative to cut-score – grade boundaries historically set

- **Workplace-based assessments**

- Formative assessments in the workplace (Norcini 2007)
- Students select cases when on placement – minimum no. each year
- Assessors (clinicians) complete level of ‘entrustability’ (Cate 2018) and narrative feedback
- App – redesigned in 2017 – no numbers – assessment for learning!
- Feedback is key - student ‘action plans’



Example of entrustability scale

Level	Descriptor
1	"I had to do" i.e., requires complete hands on guidance, did not do, or was not given the opportunity to do
2	"I had to talk them through" i.e., able to perform tasks but requires constant direction
3	"I had to prompt them from time to time" i.e., demonstrates some independence, but requires intermittent direction
4	"I needed to be in the room just in case" i.e., independence but unaware of risks and still requires supervision for safe practice
5	"I did not need to be there" i.e., complete independence, understands risks and performs safely, practice ready

Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa surgical competency operating room evaluation (O-SCORE): a tool to assess surgical competence. Acad Med. 2012;87(10);1401-7.

- **Overall MBChB decision**

- An ungraded degree – pass/fail
- But…
- There are degrees with ‘Honours’ – to reward excellence – based on graded performance across the years
- Also, access to Foundation training in the NHS is based student rank (decile) – derived again from graded assessments

- **Wider assessment trends in med. ed. assessment**

- Valuing the subjective (Hodges 2013) – nothing is objective
- Sequential models of assessment (Homer 2017) – shorter tests (with higher passing score) for the higher performers – adaptive testing – solving the ‘resit’ problem
- Programmatic assessment (Schuwirth 2017) – aka continuous assessment (?)
- Entrustable Professional Activities (EPAs – Cate 2018) – focus on level of responsibility that can be bestowed

- **UK Medical Licensing Assessment - 2020**

- To ensure a ‘common threshold for safe practice’ - GMC
- Access to the ‘register’ will require passing this in addition to medical school
- **Applied Knowledge Test** – separate test in final year run by the GMC
- **Clinical and Professional Skills Assessment** – GMC to set requirements that medical schools must comply with in their own OSCEs etc.
- Both ungraded?

- **Conclusions**

- Biggest concern is with ensuring patient safety
- Systemic pressures to grade – Honours and training
- So grading still happens – always imperfect
- Tension between ensuring competence and the need to reward excellence
- There is always progression to the next level until consultant
- Is medical education assessment/grading a (wonderful) bubble or can you learn from us?

Select references

- Cate, O. ten 2018. A primer on entrustable professional activities. *Korean Journal of Medical Education*. **30**(1), pp.1–10.
- Hays, R., Gupta, T.S. and Veitch, J. 2008. The practical value of the standard error of measurement in borderline pass/fail decisions. *Medical Education*. **42**(8), pp.810–815.
- Hodges, B. 2013. Assessment in the post-psychometric era: learning to love the subjective and collective. *Medical teacher*. **35**(7), pp.564–568.
- Homer, M., Fuller, R. and Pell, G. 2018. The benefits of sequential testing: Improved diagnostic accuracy and better outcomes for failing students. *Medical Teacher*. **40**(3), pp.275–284.
- Norcini, J. and Burch, V. 2007. Workplace-based assessment as an educational tool: AMEE Guide No. 31. *Medical Teacher*. **29**(9), pp.855–871.
- Pell, G., Fuller, R., Homer, M. and Roberts, T. 2010. How to measure the quality of the OSCE: A review of metrics - AMEE guide no. 49. *Medical Teacher*. **32**(10), pp.802–811.
- van der Vleuten, C.P.M. and Schuwirth, L.W.T. 2005. Assessing professional competence: from methods to programmes. *Medical Education*. **39**(3), pp.309–317.

Thanks

Questions?
Comments?

m.s.homer@leeds.ac.uk

@LeedsARG

Reflections of grading VTQs

- Professor Steve Higgins, Durham University
- Dr Gerard Lum, Kings College London
- Dr Matthew Homer, University of Leeds

- Chaired by Beth Black, Ofqual

ofqual

Why asking ‘why?’ is vital for strategy on grading in vocational qualifications

Tim Oates CBE

Cambridge Assessment

Why asking ‘why?’ is vital for strategy on grading in vocational qualifications

Tim Oates CBE

Group Director of ARD

December 2018

Why?

Commendation – the Ofqual research is essential

Variation in practice – deliberate, arbitrary?

Method for investigating, understanding and explaining variation

Understanding the impact of and rationale for this variation

Enhancing policy formation and qualifications development – there is a lot to play for: enhanced selection & access to learning and employment, beneficial washback into learning, lower workload for educators

Cambridge Assessment

Pass

Merit

Distinction

Pass

Distinction

Consolidating the science…

Making judgements

Aggregating judgements

Generalisability - validity

Decomposition

Recomposition

Do things survive recomposition?

Artefacts and anomalies

Approximately content with the components and discontent with the overall product

Williamson J 2018 ***Characteristics, uses and rationales of mark-based and grade-based assessment*** Cambridge Assessment

What criterion lies behind the grade

What norm lies behind the criterion

What language and processes mean that this is dependable

Level 4A

What ‘levels’ meant?

As a guide, here’s what national curriculum level the Government suggested a child should achieve by the end of each school year:

Year 1	Level 1b
Year 2	Level 2a-c
Year 3	Level 2a-3b
Year 4	Level 3
Year 5	Level 3b-4c
Year 6	Level 4

By the end of Year 6, approximately 75 per cent of children will achieve a Level 4; the top 10 per cent will achieve a Level 5, and the ‘exceptional’ top one per cent, a Level 6. Children who achieve a Level 4 are expected to go on to pass five or more GCSEs at grade A*-C.

Lucy Dimbylow **The School Run**

https://www.theschoolrun.com/what-are-national-curriculum-levels

A*

A*-C

A*-C

9-1

A*-C

9-1

1-9

99

Establishing the meaning of the score – STA

Range of scaled scores

The range of scaled scores available for each test is the same as set in 2016 and is intended to stay the same in future years. The lowest scaled score that can be awarded on a KS1 test is 85. The highest scaled score is 115.

Pupils scoring at least 100 will have met the expected standard in the test.

A pupil awarded a score of 99 or fewer has not met the expected standard in the test.

Pupils need a minimum raw score before they can be awarded the lowest scaled score. Pupils who do not achieve the lowest scaled score on the test have not demonstrated sufficient understanding of the KS1 curriculum in the subject. You should award these pupils an N for the test. It is likely that these pupils should be teacher assessed using the pre-key stage standards.

HE essay mark

A+/B++?

Intense ambiguity

Working at Grade 5

Necessary theory and consideration

Technical quality of the grades

Culture

Use of Grades – but ‘meaning’ conveys something more

Meaning

Interpretation

Use in accountability

Washback into learning

Impact

Individual identity

Crystallised social structures

Knowledge exchange between discrete communities

Lack of transaction of knowledge and research between General Qualifications and Vocational Qualifications

Encouraging insights from technical and vocational area Influencing thinking in General Qualifications

GCSE Science assessment model – profoundly influenced by VET

GCSE Science - radical model with significant education benefit

The nature of learning in practical activity - informed by concepts of 'competence' in respect of professional and technical education

HE interests regarding threshold measures

Learning from practical work - failure and situated cognition

Looking at grading approaches with a wider lens

Hidden/wider structures and impacts on grading

Looking at individual qualifications or components in the vocational domain can be misleading - interesting mixed models

Component scoring and grading

‘Long term modular certification’ - accumulating credit through components

Interesting aspects of the landscape

Very odd ideas regarding ‘progress’ and indicative grades

T Levels - potentially located/stuck in the middle

Issues

From my 2008 presentation… still relevant ten years on?

A few remaining issues:

1 why throw away information?
2 HEIs are turning to finer-grained information – module grades, UMS
3 dealing with scores – will scores be over-interpreted, despite our best warnings?
4 can profiles of attainment be used in progression and other key arrangements?
5 will conversion to UMS still be trammelled with undesirable artefacts?
6 can we handle probabilistic data?

Issues

The continued issue of ‘right in general’ in tension with ‘right for each individual’

Where error resides in the reporting - concentrating error around the boundaries

Will any grading system pass intense public scrutiny regarding ‘fairness’

The approximate nature of ‘mastery’

‘A-like’ and ‘G-like’ individuals - the ontology of grades and assessment objectives

And finally … the future

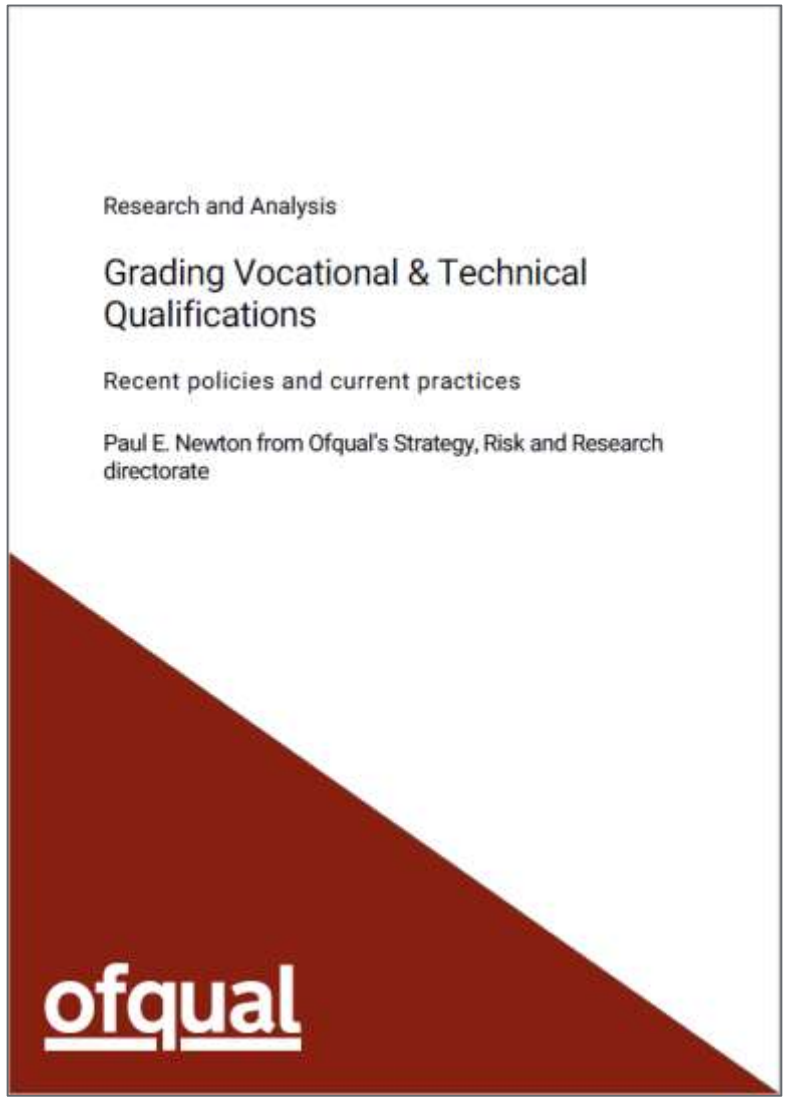
Closing comments

Paul Newton

Recap

- Small scale research covering:
  - 18 qualifications at level 1 – level 4
  - 15 awarding organisations many of whom are here today
- Literature review Australia
- Development of a statement of principle for grading in vocational assessments …

A very high-level statement of principle



1. **The grounds for differentiating between candidates, via grades, must be defensible; that is, sufficiently meaningful and sufficiently useful, when judged in relation to a profile of purposes.**
2. **The grading process must be sufficiently accurate.**
3. **The benefits from implementing the grading process must, on balance, outweigh its costs.**

What should AOs / EPAOs do next?

- Recommendations: The report does not have recommendations for Ofqual or AOs. We do not expect any AOs or EPAOs to change their approach having read this report; instead we are continuing a conversation with the academic and AO community about practice.
- AOs: we encourage AOs to think about the rationale for grading within their qualifications and take part in the conversation with Ofqual, the regulated community, academics and stakeholders over the next year.
- EPAOs: approaches to grading at the level covered in the research are presented in the Assessment Plan, written by a Trailblazer group. EPAOs should continue to work in accordance with the parameters and guidance set by the Assessment Plan. Ofqual is not asking AOs to review their assessments. Trailblazers may wish to consider this research as they develop an Assessment Plan.

What next for the conversation?

- Smaller-scale events
  - e.g. Webinar with FAB January 2019
- Continue to monitor issues related to
  - current qualifications and assessments
  - future qualifications and assessments
- Further discussions with AOs, centres and stakeholders
- Further research and analysis both within Ofqual and the wider academic or research community

To contact us

- Thank you for participating

- We look forward to continuing the conversation with you

  - AOs and Ofqual EPAOs please use the Portal **message** function to Contact Ofqual

  - All other stakeholders please contact Sian.Sankey@ofqual.gov.uk

- **Feedback: smartsurvey.co.uk/s/Ofqual11Dec**