Write or wrong?

Rethinking essays: a comparative study of very short answer vs multiple choice questions

University of Central Lancashire
UCLan

Dr Harry Potter

harry.potter@manchester.ac.uk

Lecturer in Neuroscience

Division of Neuroscience, SBS, FBMH

MANCHESTER 1824
The University of Manchester

Potter & McLachlan 2025

Medical Teacher

MEDICAL TEACHER
https://doi.org/10.1080/0142159X.2025.2496382

MEDICAL TEACHER

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Assessing medical knowledge: A 3-year comparative study of very short answer vs. multiple choice questions

Harry G. Potter[a,b] and John C. McLachlan[b]

[a]Division of Neuroscience, School of Biological Sciences, Faculty of Biology, Medicine and Health, Manchester Academic Health Science Centre, University of Manchester, Manchester, United Kingdom; [b]School of Medicine and Dentistry, University of Central Lancashire, Preston, United Kingdom

ABSTRACT

Purpose: Assessment design significantly influences evaluation of student learning. Multiple choice questions (MCQ) and very short answer questions (VSAQ) are commonly used assessment formats, especially in high-stakes settings like medical education. MCQs are favoured for efficiency, coverage, and reliability but may lack depth in assessing critical thinking. VSAQs require students to generate responses, potentially enhancing depth, but posing challenges in consistency and subjective interpretation.

Methods: Data from parallel MCQ/VSAQ exams over three years was collected. Summary statistics for each exam (marks, time, and discrimination index; DI) and the effect of year and question characteristics were analysed.

Results: VSAQs were associated with lower marks ($p < 0.001$), longer time ($p < 0.001$), and higher DI ($p < 0.001$). Question characteristics (e.g. basic science or clinical stems) significantly affected the mark, time, and DI, changing across years, but not interacting with question format.

Conclusion: While MCQs resulted in higher marks, VSAQs provided higher discrimination of student performance. Response options in MCQs likely enhance recall, however real-world settings also offer contextual cues. Question characteristics affect student performance independently of format, likely due to differences in cohort career progression. Future research should investigate predictive validity and standard setting of VSAQs in a basic science context.

ARTICLE HISTORY
Received 22 November 2024
Accepted 17 April 2025

KEYWORDS
Assessment; Multiple choice questions; Very short answer questions

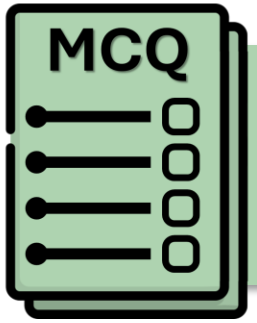
Abstract of Potter & McLachlan 2025 in Medical Teacher.

- No conflicts to declare.

FULL PAPER

Introduction

- HE providers often use **high stakes essays**.
- Assess **depth** well – may lack breadth & widen **attainment gaps**[1].
- Students report **unfavorability** & desire **varied assessments**[2].



✓ **Efficiency**, coverage/breadth, reliability.

⊗ Cueing, **depth**.

✓ **Recall, depth,** & authenticity.

⊗ Utility less well understood.

VSAQs are:

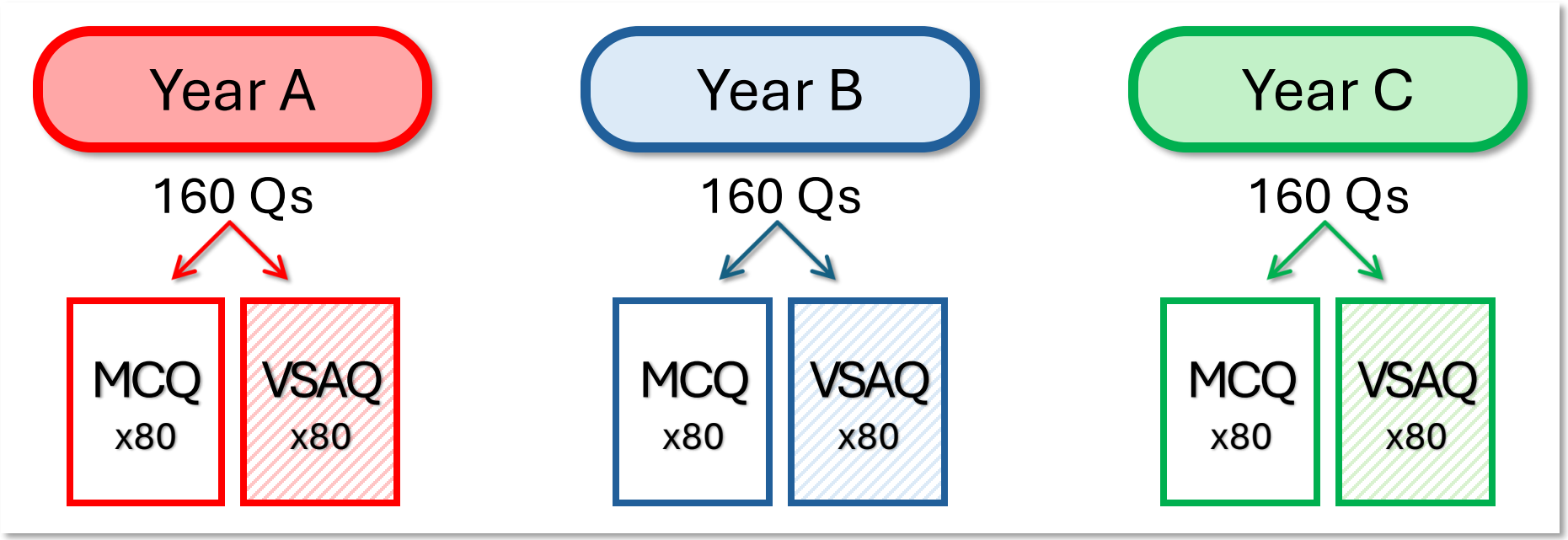
- Better discriminators[3-5]
- Harder[3,5]
- Authentic[3]
- Acceptable[3]

FULL PAPER

[1]Canal & Child 2025; [2]EAR 1.0; [3]Sam et al. 2018; [4]Sam et al. 2019; [5]Neumann et al. 2021

Methods (1) – teaching & assessment

- 40 credit, level 5 immunology unit.
- 3 years, (largely) same delivery.
- Compulsory, parallel MCQ and VSAQ exams.

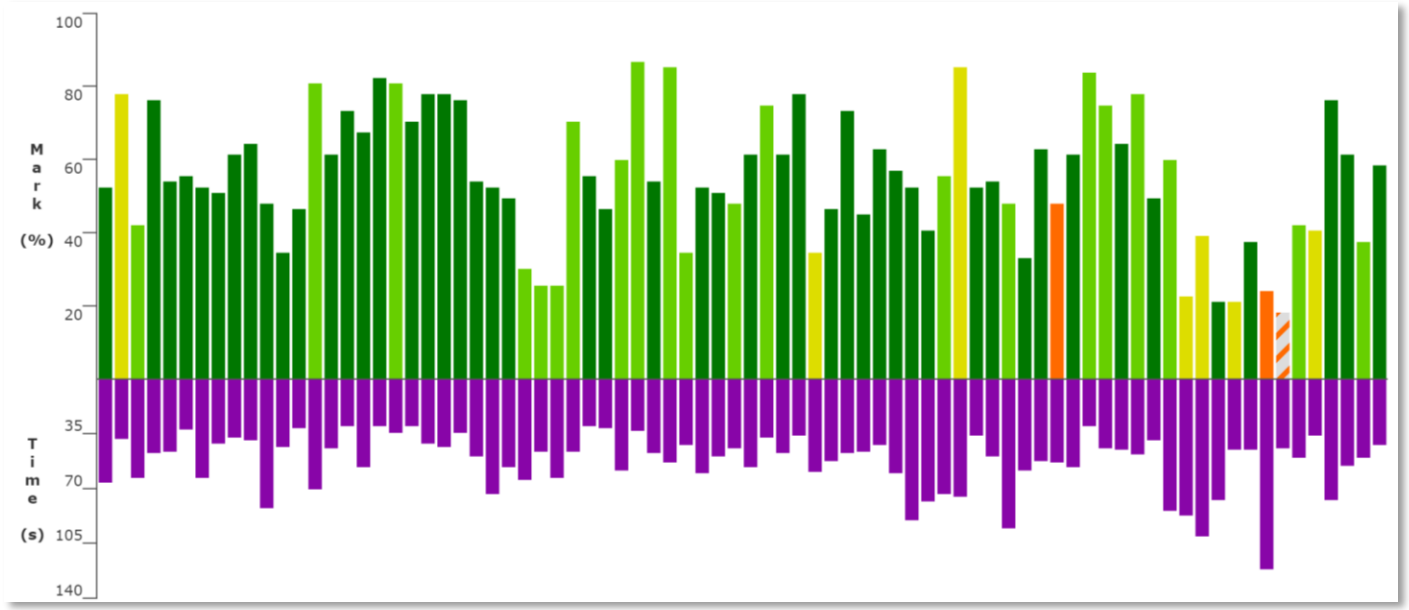


Assessment design with 80 MCQ and 80 VSAQ questions per year.

FULL PAPER

Methods (2) – outcome measures

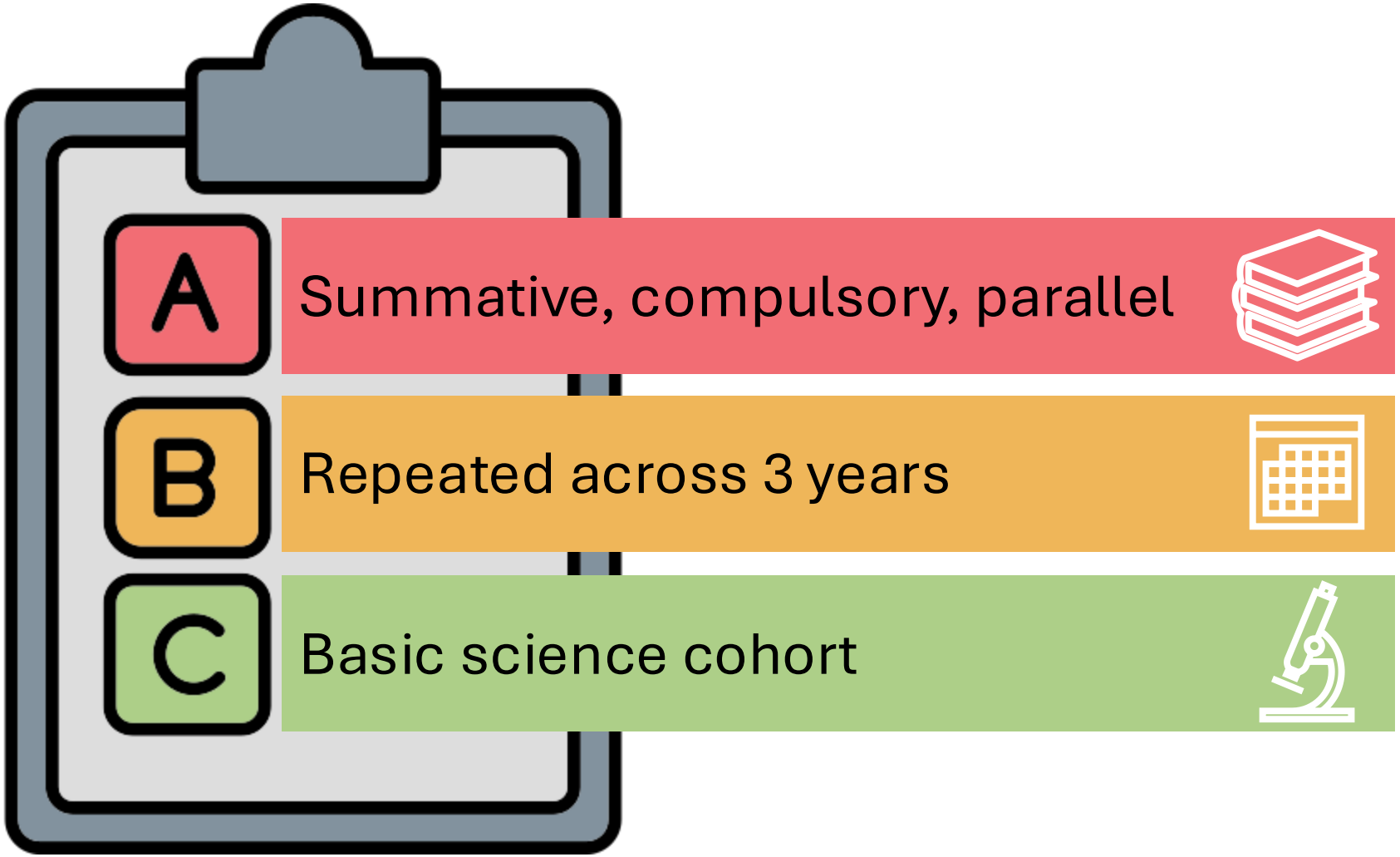
- **Students:** marks.
- **Questions:** time, discrimination.
- **Characteristics**: clinical stems, images, concept.



Representative exam data output from Maxinity.

FULL PAPER

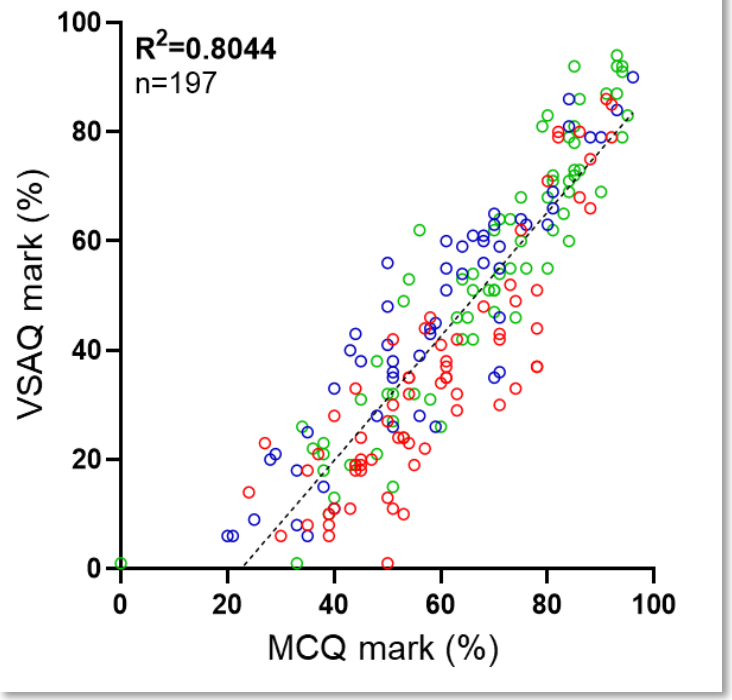
Significance



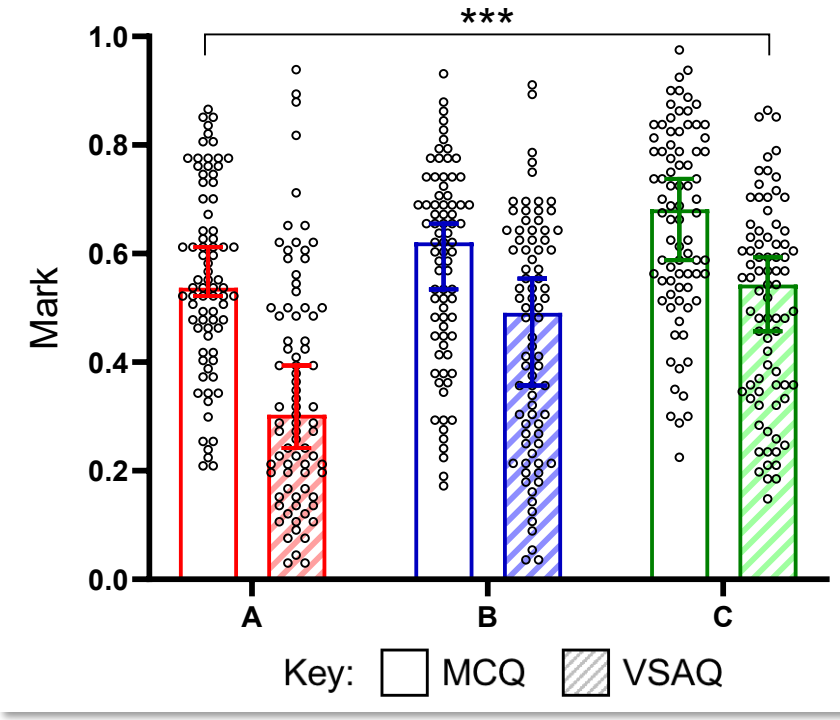
FULL PAPER

Results (1) – how do MCQs and VSAQs compare?

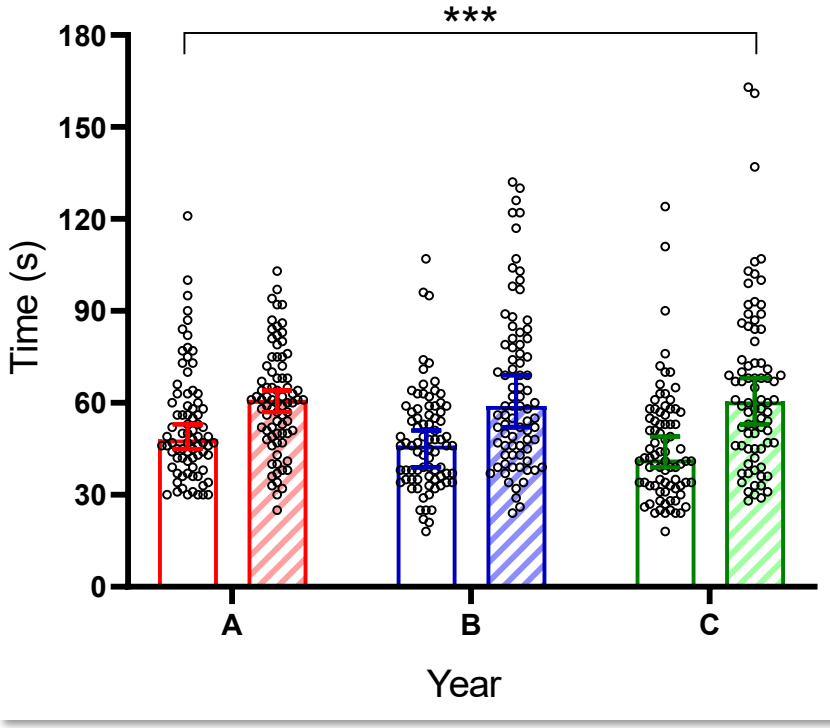
i) MCQ vs VSAQ marks



ii) Mark per question



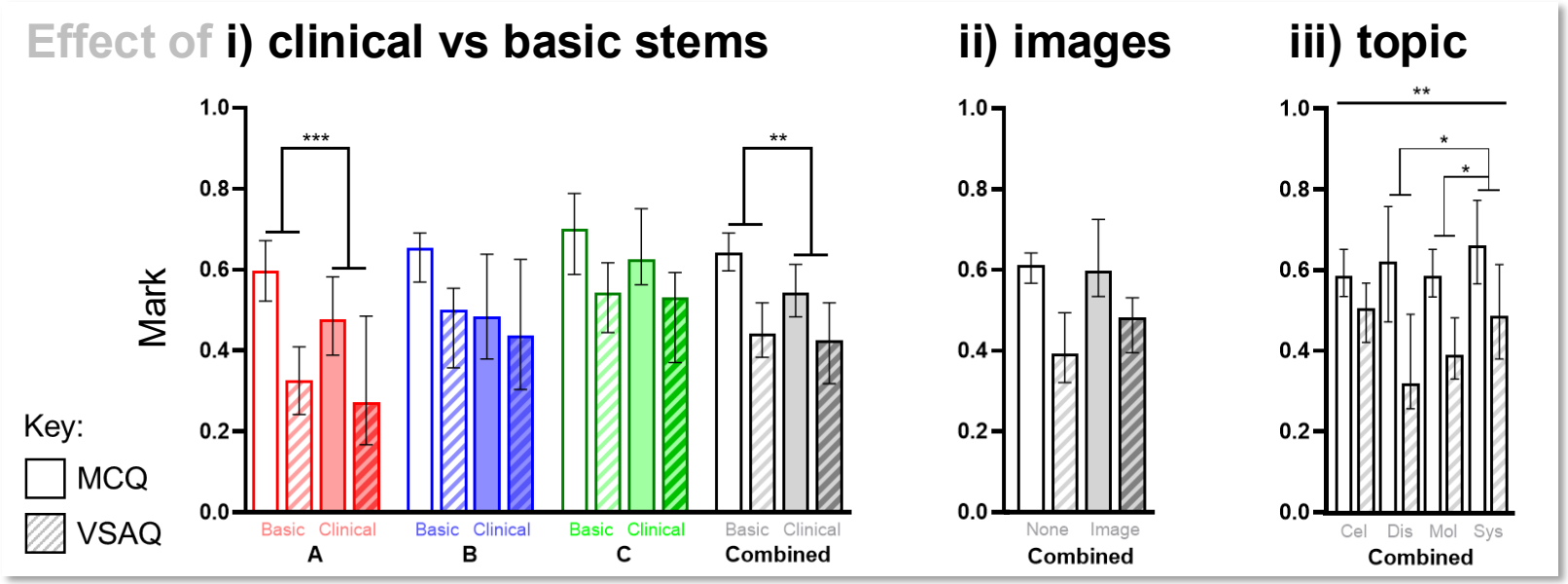
iii) Time per question

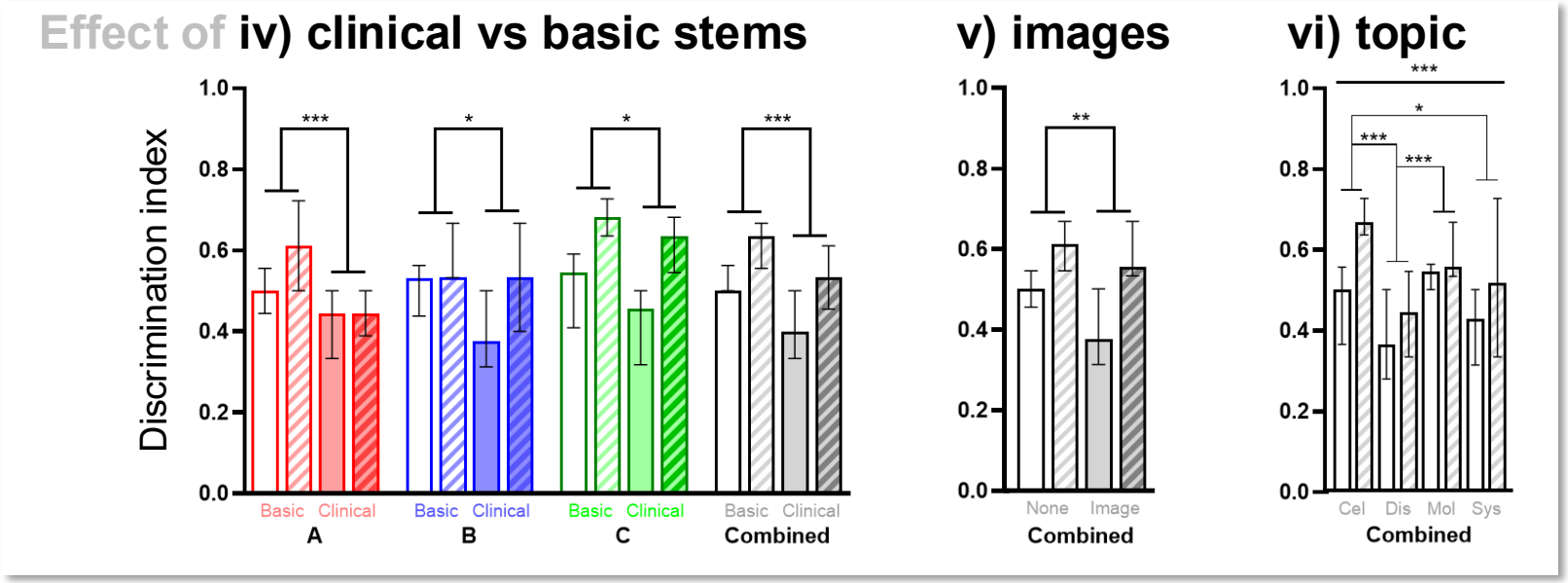


i) Correlation of MCQ and VSAQ marks across years A-C (n=56-73 students per year). **ii)** Mean mark per question across MCQ (open bars) and VSAQ (hatched bars) exams in years A-C (n=78-80 questions per exam). **iii)** Mean time per question across MCQ (open bars) and VSAQ (hatched bars) exams in years A-C (n=78-80 questions per exam). Key: ***p<0.001.

FULL PAPER

Results (2) – effect of question characteristics





- Q characteristics affect mark/discrimination (& time) **independent of format**.
- **Temporal** patterns in VSAQ use.

Key: Cel, cellular; Dis, disease; Mol, molecular; Sys, systems. *p<0.05; **p<0.01; ***p<0.001.

FULL PAPER

Discussion & Future Research

MCQ/VSAQ marks are highly correlated – **construct**

VSAQs are **harder**, take longer, and **better discriminate**

Q characteristics affect performance **independent of MCQ/VSAQ**

Supports use in a **varied/inclusive assessment design** in basic sciences.

- **Practical implementation** – temporal trends, IT support, staff training[6].
- **Standard setting**[7] – recalibration of staff/student expectations.
- **Predictive validity**[8,9] – do VSAQs predict later competencies?

[6]Tarrant & Ware 2012; [7]Sam et al. 2022; [8]McManus et al. 2013; [9]Wakeford et al. 2018

FULL PAPER

Thank you, Qs, & future work

Dr Harry Potter

harry.potter@manchester.ac.uk

Lecturer in Neuroscience

Division of Neuroscience, SBS, FBMH

Acknowledgements

Prof John McLachlan

UCLan Burnley assessment team

Bridget Gardner

Sarah Bartlett

Charlotte Taylor

University of Central Lancashire UCLan

References

[1]Canal MM & Child R (2025). *Assess. Eval. High.* 1:1-4.

[2]Enhancing Assessment Review, UoM 2025.

[3]Sam AH et al. (2018). *Med. Educ.* 52:447-455.

[4]Sam AH et al. (2019). *BMJ Open* 9:e032550.

[5]Neumann J et al. (2021). *Educ. Res Int.* 2021:1–10.

[6]Tarrant M & Ware J (2012). *Nurse Educ. Today* 32:e19–e24.

[7]Sam AH et al. (2022). *BMC Med. Educ.* 22:640.

[8]McManus IC et al. (2013). *BMC Med.* 11:244.

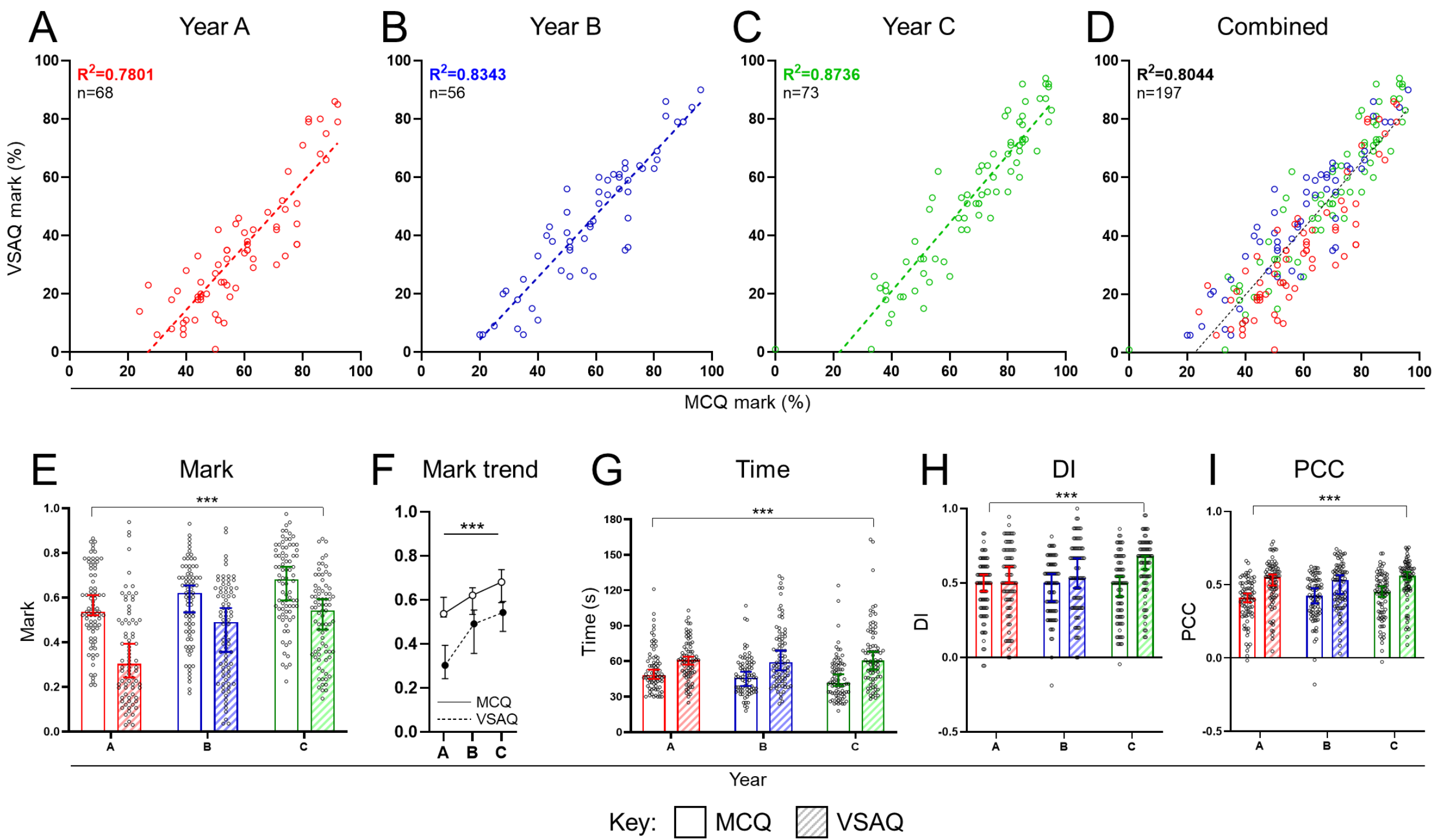
[9]Wakeford R et al. (2018). *BMC Med.* 16:230.

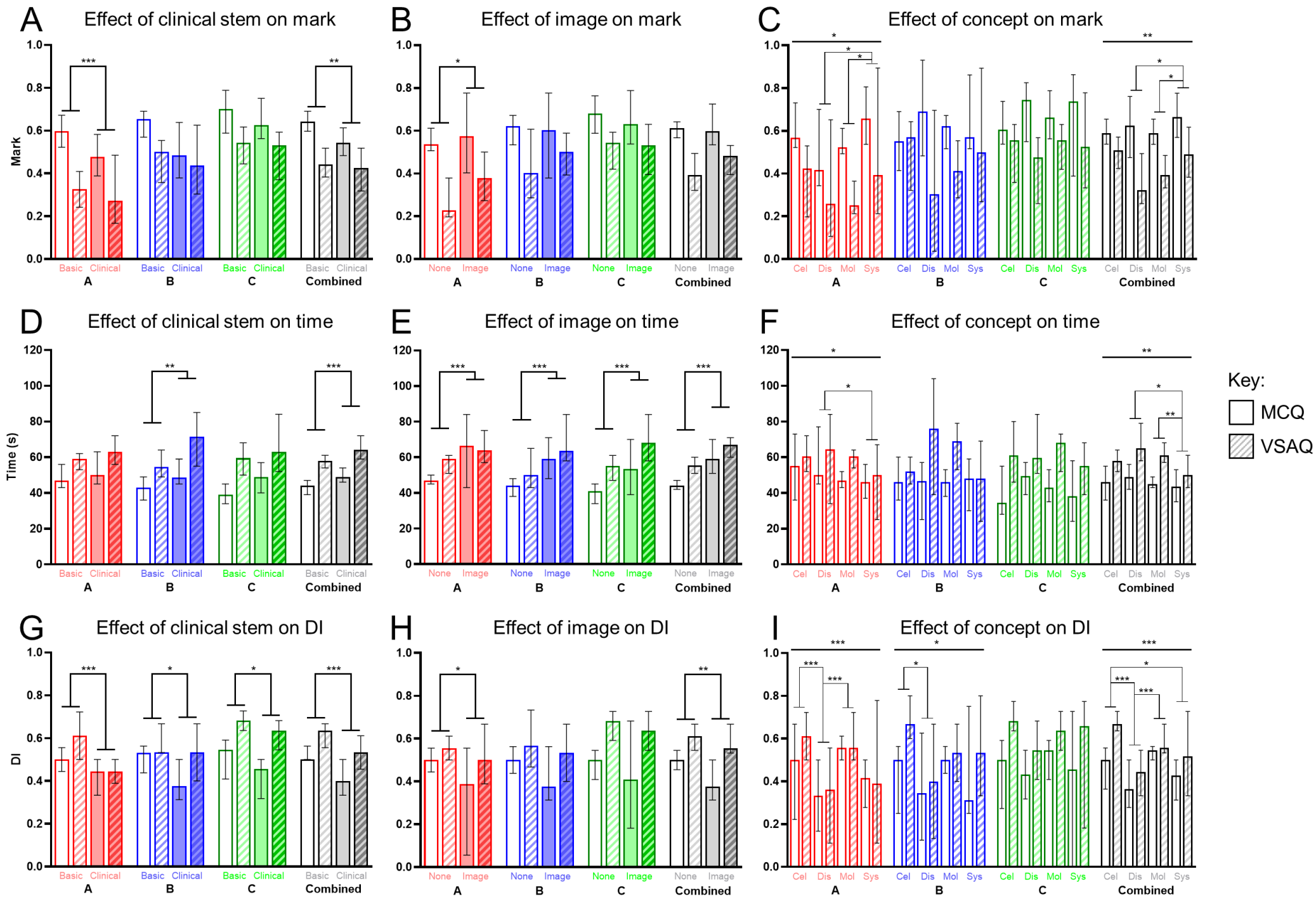
FULL PAPER

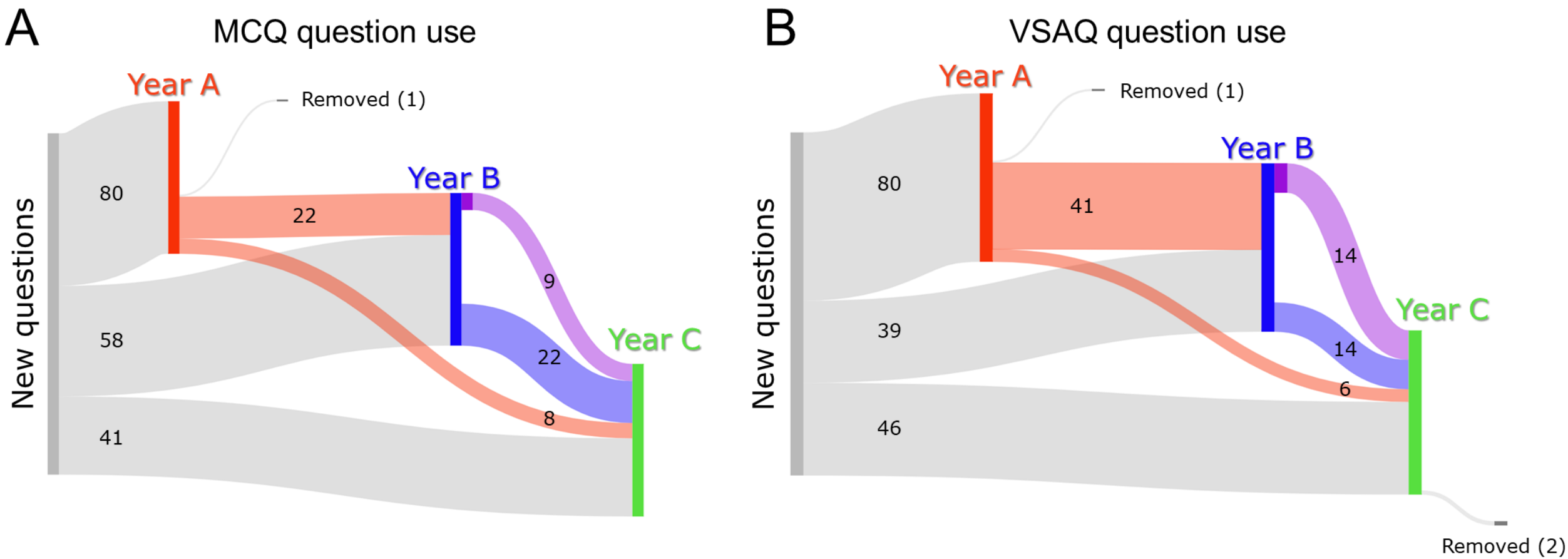
Table 1. Summary data for MCQ and VSAQ assessments across years A–C.

	Year A		Year B		Year C	
	MCQ	VSAQ	MCQ	VSAQ	MCQ	VSAQ
Number of students	68	68	56	56	73	73
Number of questions	79	79	80	80	80	78
Minimum mark (%)	21.5	1.3	20.0	6.3	27.5	1.3
Maximum mark (%)	91.1	86.1	96.3	90.0	95.0	93.6
Mean mark (%)	55.8	35.1	57.8	45.1	65.9	50.6
Standard deviation (%)	18.3	21.8	18.7	22.0	18.6	24.4
Cut score (%)	35.2	40.0	37.5	40.0	40.0	40.0
% Who failed (at same mark)	11.9 (20.9)	66.7	15.5 (17.9)	41.1	12.5 (12.5)	37.0
Discrimination index (DI)	0.452	0.523	0.461	0.547	0.462	0.615
Cronbach's alpha	0.929	0.964	0.935	0.960	0.942	0.967
Standard error (%)	4.9	4.2	4.7	4.4	4.5	4.5
Difference in means		20.7		12.7		15.3
Actual mean difference		20.9		13.0		14.7
Slope		1.0508		1.0673		1.2109
Coefficient of determination		0.7801		0.8343		0.8736
Correlation		0.8832		0.9134		0.9347
Disattenuated correlation		0.9862		1.0176		1.0261

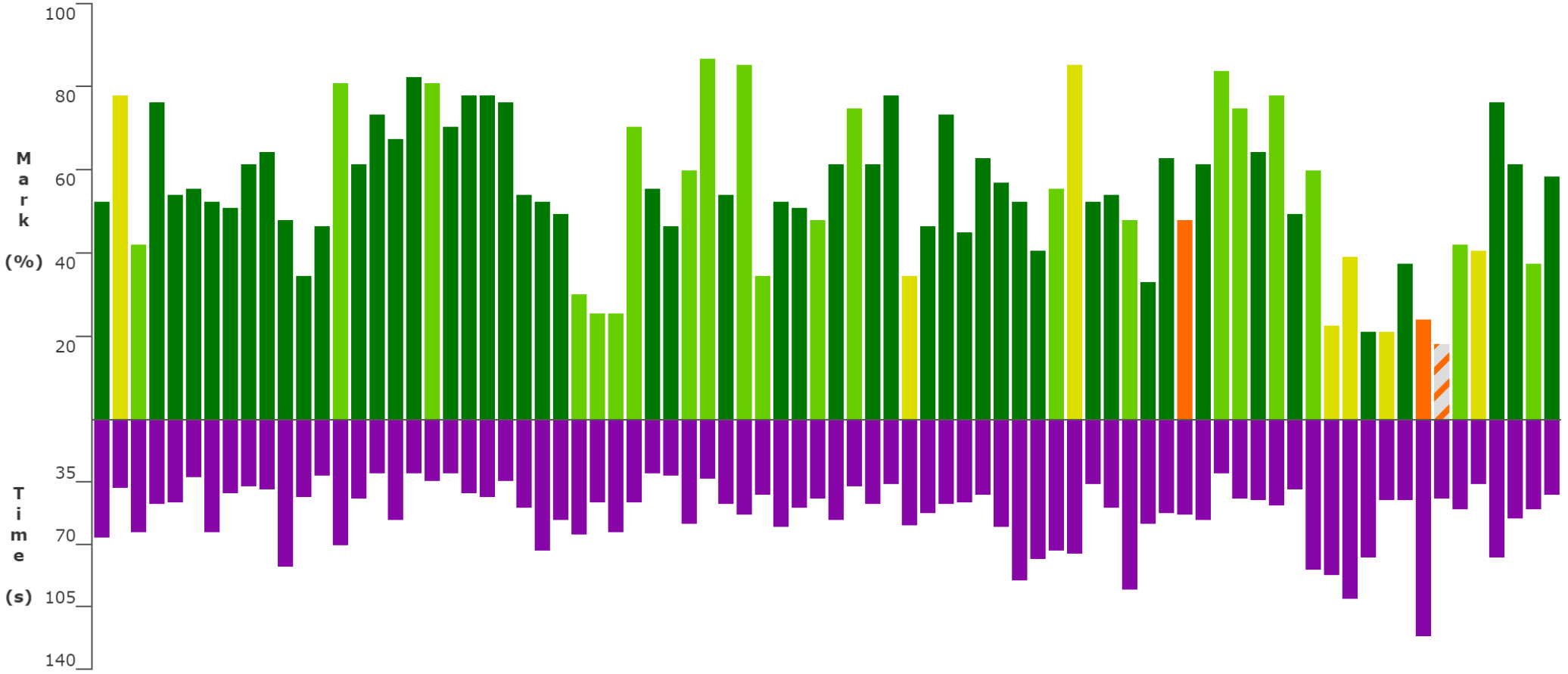
Discrimination index, DI; multiple choice question, MCQ; very short answer question (VSAQ).


A
Year A
R2=0.7801
n=68
B
Year B
R2=0.8343
n=56
C
Year C
R2=0.8736
n=73
D
Combined
R2=0.8044
n=197
VSAQ mark (%)
MCQ mark (%)
E
Mark
F
Mark trend
G
Time
H
DI
I
PCC
***
MCQ
VSAQ
Year
Key: MCQ VSAQ


A
Effect of clinical stem on mark
B
Effect of image on mark
C
Effect of concept on mark
D
Effect of clinical stem on time
E
Effect of image on time
F
Effect of concept on time
G
Effect of clinical stem on DI
H
Effect of image on DI
I
Effect of concept on DI
Mark
Time (s)
DI
Basic Clinical
None Image
Cel Dis Mol Sys
A
B
C
Combined
Key:
MCQ
VSAQ


A
MCQ question use
Year A
Removed (1)
Year B
New questions
80
58
41
22
9
22
8
Year C
B
VSAQ question use
Year A
Removed (1)
Year B
80
39
46
41
14
14
6
Year C
Removed (2)

Mean mark for each contributing item



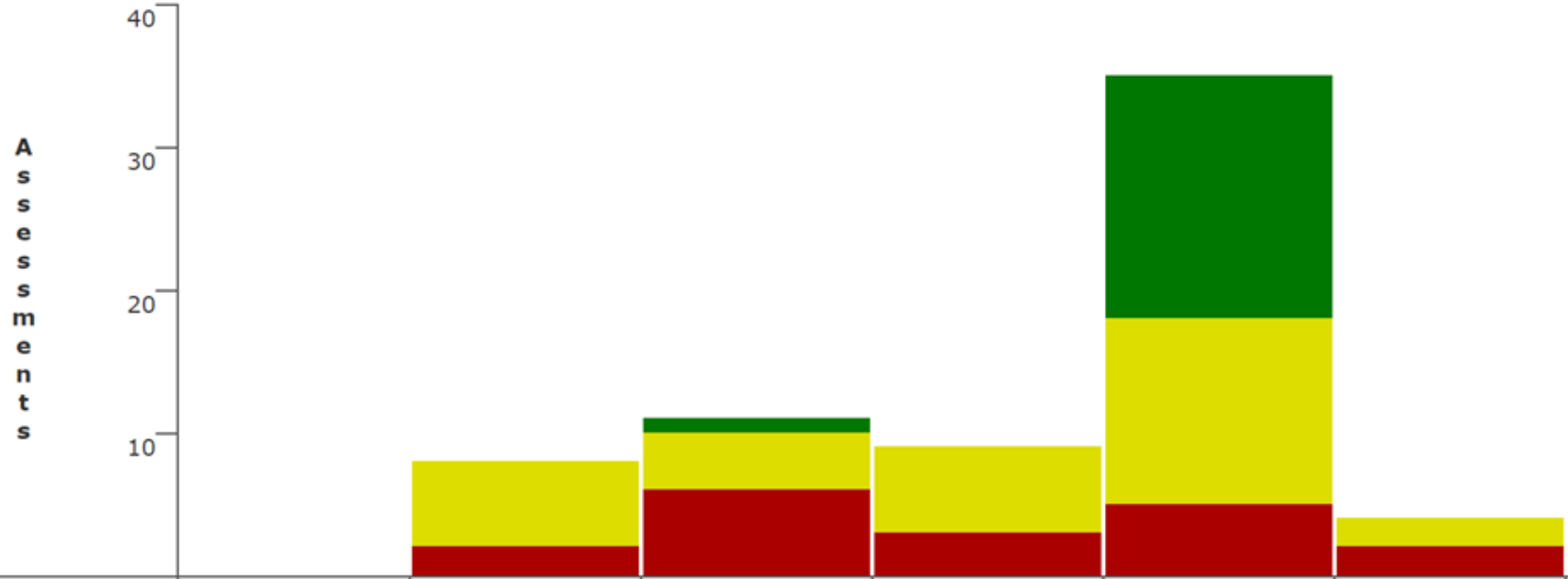
Time per question per available mark

Key

- 0.4 ≤ DI
- 0.2 ≤ DI < 0.4
- 0.0 ≤ DI < 0.2
- -0.2 ≤ DI < 0.0
- DI < -0.2
- Disabled item

	Disabled items excluded	All items
Maximum mark	79	80
Mean mark	44.09 55.8%	44.27 55.3%
Exam DI	0.459	0.452
Standard deviation	14.42 18.3%	
Std err of measure	3.84 4.9%	
Cronbach's alpha	0.929 item based	

No. of assessments per option



Option	No answer	A	B	C	D	E	Overall	Mean mark
Option mark		✗	✗	✗	✓	✗	1	
Discrimination		-0.111	-0.278	-0.167	0.667	-0.111	0.667	
Correlation		-0.165	-0.196	-0.267	0.514	-0.168	0.514	
Upper group	0	0	1	0	17	0	18	0.944
Middle group	0	6	4	6	13	2	31	0.419
Lower group	0	2	6	3	5	2	18	0.278
Total	0	8	11	9	35	4	67	0.522

Key

- Upper group
- Middle group
- Lower group

Horst PKI 0.358

Standard setting BMedSci Year 2 context

BMedSci Year 2 Ebel - Ebel: Set to Hard, Supplementary

Criteria to be considered when developing as assessment strategy.

Criterion	Description	Reference
Construct validity	The assessment measures the intended knowledge or behavior	Gulikers 2004, Schuewirth 2004, Norcini 2011
Reproducibility	The assessment would yield the same results if repeated under similar circumstances	Norcini, 2011
Equivalence	The assessment yields same or similar scores when given across institutions	Norcini, 2011
Feasibility	The assessment is practical and realistic	Norcini, 2011
Authenticity	The assessment requires learners to apply the same knowledge, skills, and attitudes needed in real-life professional situations	Gulikers, 2004
Acceptability	All stakeholders (e.g., learners, instructors, administrators) agree that assessment process and results are credible	Norcini, 2011
Educational effect	The assessment prompts educationally beneficial preparation by learners	Norcini, 2011 Hift, 2014
Testing effect	Information retrieval improves later recall	Roediger 2011
• Generation effect	Assessment of content improves later recall	Taconnat, 2008 Roediger 2011
• Cueing effect	Answer recognition from a list of choices may overestimate student knowledge	Veloski 1999 Sam, 2018
Catalytic effect	Assessment results and feedback contributes to future learning	Norcini, 2011 Hift, 2014

Bird et al. 2019
Med Educ Online

[1]Canal MM & Child R (2025). Impact of the type of assessment on awarding gaps in bioscience undergraduate degrees. *Assess. Eval. High.* 1:1-4.

[2]Enhancing Assessment Review, UoM 2025.

[3]Sam AH et al. (2018). Very-short answer questions: reliability, discrimination and acceptability. *Med. Educ.* 52:447-455.

[4]Sam AH et al. (2019). Comparing single-best answer and very-short-answer questions for the assessment of applied medical knowledge in 20 UK medical schools: cross-sectional study. *BMJ Open* 9:e032550.

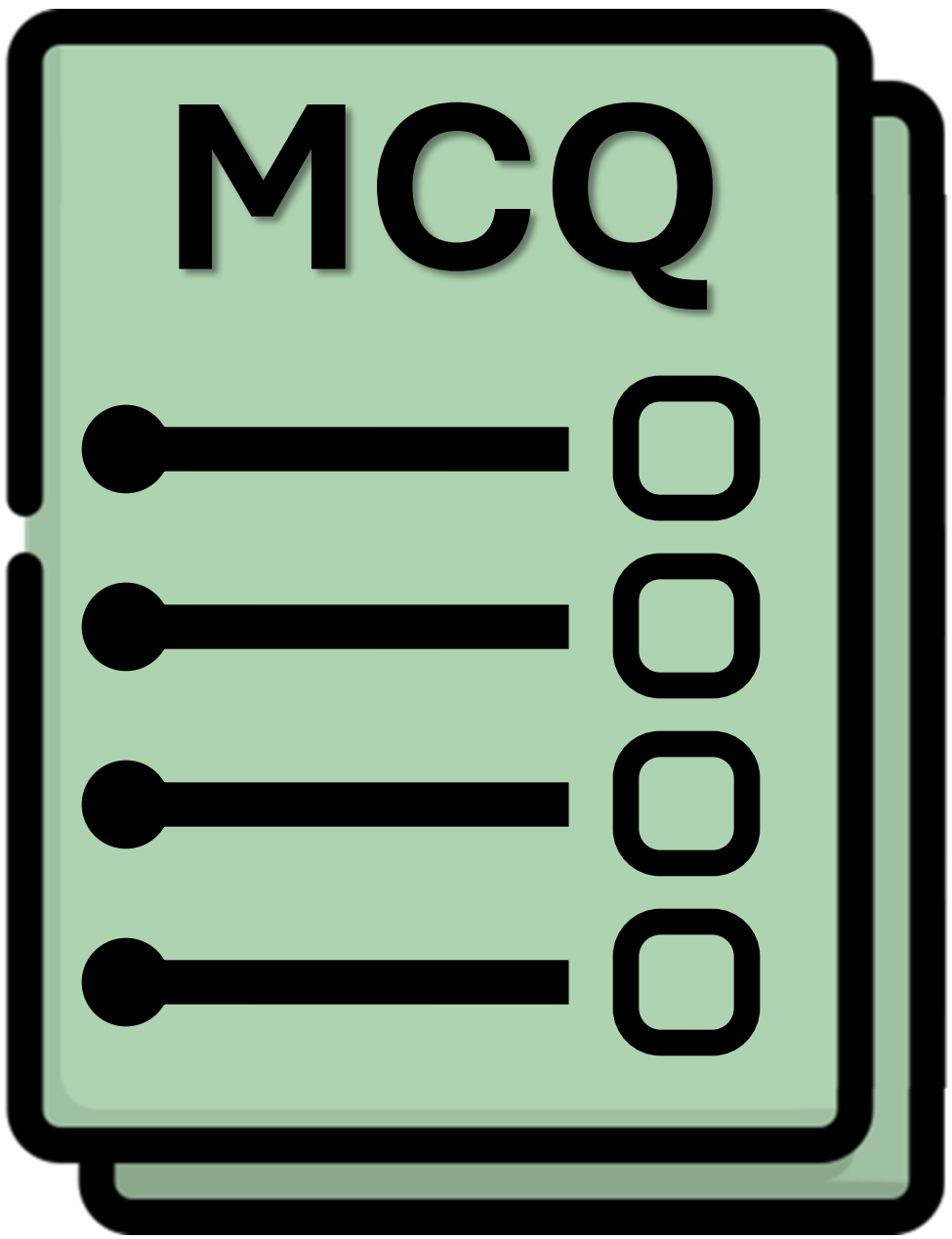
[5]Neumann J et al. (2021). Comparison of online tests of very short answer versus single best answers for medical students in a pharmacology course over one year. *Educ. Res Int.* 2021:1–10.

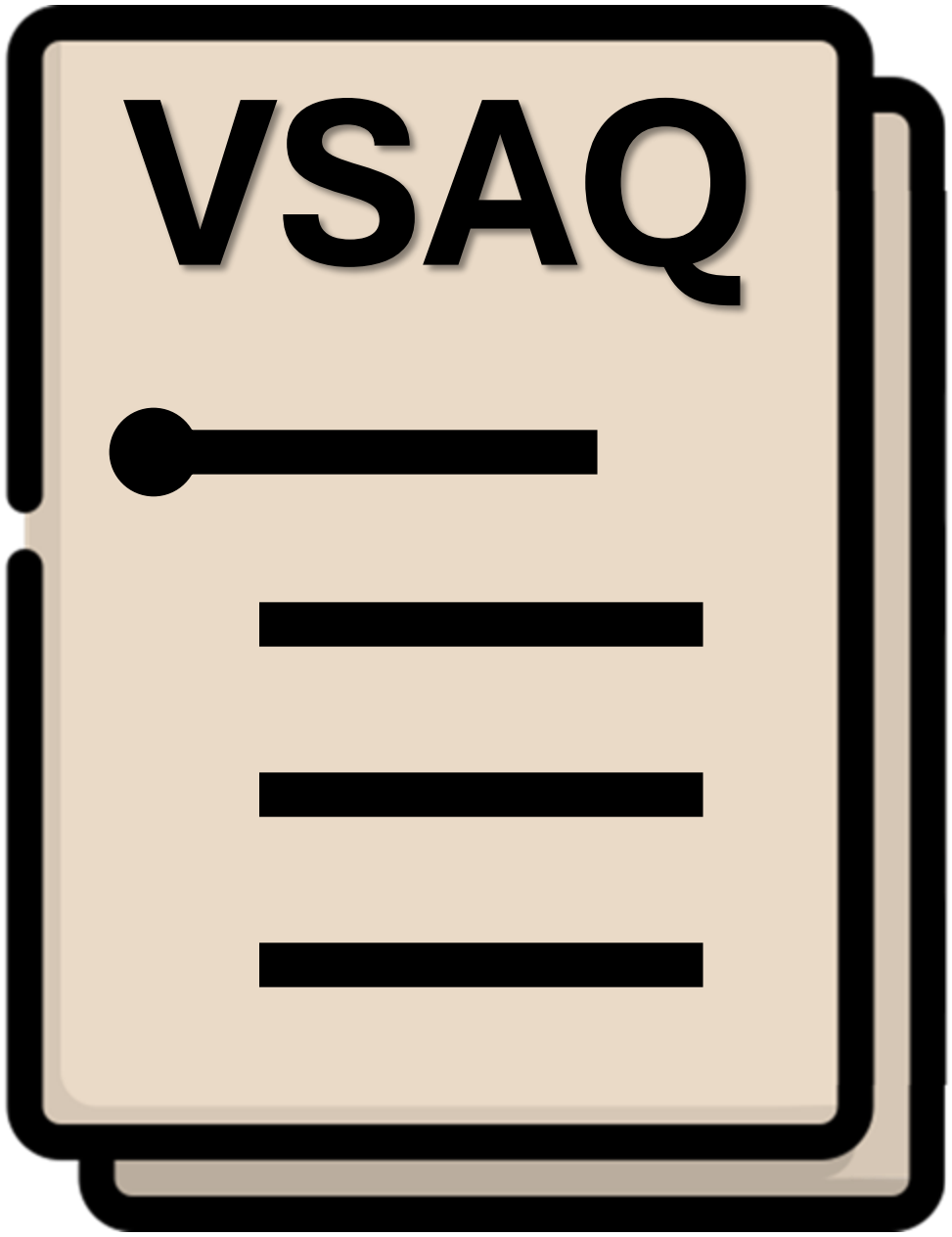
[6]Tarrant M & Ware J (2012). A framework for improving the quality of multiple-choice assessments. *Nurse Educ. Today* 32:e19–e24.

[7]Sam AH et al. (2022). Standard setting very short answer questions (VSAQs) relative to single best answer questions (SBAQs): does having access to the answers make a difference? *BMC Med. Educ.* 22:640.

[8]McManus IC et al. (2013). The UKCAT-12 study: educational attainment, aptitude test performance, demographic and socio-economic contextual factors as predictors of first year outcome in a cross-sectional collaborative study of 12 UK medical schools. *BMC Med.* 11:244.

[9]Wakeford R et al. (2018). Fitness to practise sanctions in UK doctors are predicted by poor performance at MRCGP and MRCP(UK) assessments: data linkage study. *BMC Med.* 16:230.


MCQ


VSAQ