Pearson

The challenges of introducing and conducting a high-stakes tertiary admissions test for school-leavers

UCAT
UNIVERSITY CLINICAL APTITUDE TEST
ANZ

Dr Edward Li
Pearson Professional Assessments

A/Prof Annette Mercer
Monash University

Copyright 2026 Pearson. All rights reserved.

Agenda

Selection of medical students
Aptitude testing

Geographical challenges
Language and culture

Technical and security challenges
The test population

Construct validation

Score comparability

Measurement quality

Copyright 2026 Pearson. All rights reserved.

Selection of medical students
Aptitude testing

UCAT
UNIVERSITY CLINICAL APTITUDE TEST
ANZ

Copyright 2026 Pearson. All rights reserved.

Selection of medical students

UCAT ANZ

Selection is high-stakes

- The number of applicants is significantly greater than the number of places available
- Selection usually involves three components:
- Academic score (high)
- Aptitude test
- Interview (MMI)
- The three components are used in different ways by the universities to make their final offers.

UCAT ANZ

- An aptitude test used by a consortium of 18 universities in Australia and New Zealand.
- The test consists of three cognitive sections: Verbal Reasoning, Decision Making and Quantitative Reasoning, plus a non-cognitive section called a Situational Judgement Test.
- The test is multi-choice, computer-based and administered in test centres.
- It is strictly timed in sections.
- **Most Australian candidates are school-leavers (17 – 18 years old).**

Copyright 2026 Pearson. All rights reserved.

Challenges:

Geographical
Language and context

UCAT UNIVERSITY CLINICAL APTITUDE TEST ANZ

Copyright 2026 Pearson. All rights reserved.

The challenges

Language and context

- UCAT ANZ is conducted under license from UCAT UK.
- Early versions of the test were UK-centric – language, item context, units of measurements……
- Due to the scales set up for long-term measurement, the transition to a more 'global' version took several years.
- Current test versions have a range of different units of measurement, currencies, more universal contexts etc.

Geographical challenges

- Australia is vast continent with a relatively small population.
- There is concentration of population around the coast.
- Equity and diversity values held by the universities and reinforced by government policies mean that we need to cater for candidates who live in rural and remote areas.
- As a result, Pearson set up a number of temporary test centres in rural areas. These centres operate for 1 to 3 days to cater for local demand.
- This is an expensive but necessary exercise.

Outcomes

- Australia is a multicultural society.
- Test items need to be accessible to a wide range of candidates from different backgrounds.
- It is important to provide test conditions for rural and remote candidates which are as close as possible to those experienced by urban candidates.
- We also provide practice tests and test strategies so that candidates who live rurally or who don't have the financial resources are able to undertake preparation.

Copyright 2026 Pearson. All rights reserved.

External challenges

UCAT UNIVERSITY CLINICAL APTITUDE TEST ANZ

Copyright 2026 Pearson. All rights reserved.

External challenges

Environmental etc

- Unexpected disruptions – weather, noise distraction, building alarms.
- Preparation providers: items provided for practice are not necessarily representative of the real items.
- Candidates are given an unrealistic idea of their potential achievement. Often believe that their real test result must be wrong!

Technical

- Inevitable problems with computers, software, internet.
- Candidates who 'do the wrong thing' e.g. altering PDF results sheets.
- The prevalence of AI and other technological advances may pose threats that we cannot anticipate.
- **This is the X-factor!**

Outcomes

- Many of these challenges must be managed as well as possible, we have very little control over them.
- The potential for security issues is greater than ever.
- We need to be alert to the possibilities and deal with issues that arise in the most straightforward and equitable ways that we can.
- The nature of this cohort means that some don't necessarily cope well with disruption.

Copyright 2026 Pearson. All rights reserved.

The test population

UCAT UNIVERSITY CLINICAL APTITUDE TEST ANZ

Copyright 2026 Pearson. All rights reserved.

Our candidates

Their characteristics

- They are young
- They are smart
- They are a product of their generation, i.e. a level of entitlement.

Issues arising

- Under 16s
- Suitable and current photo ID
- Many are not independent. They are accustomed to their parents/school organising or advocating for them.
- Parents are often high-level professional/business or academic people.
- Their methods of communication/interaction with the world are different from those of other generations.

- The test centre model of delivery has proved to be the most reliable. During COVID we needed to use remote proctoring and that was very challenging for this group.
- There is a certain level of preparation needed, which many of them find difficult.
- There are rules to be followed.

- The UCAT ANZ is the first hurdle in admission to a medical course. It is high-stakes for the candidates and for the future medical workforce.

Copyright 2026 Pearson. All rights reserved.

Outcomes

- Acceptable photo ID is a challenge for some. No driver's licence, no passport. As a form of alert, we ask them to specify which ID they will use.
- Parents often speak on their behalf. We have to ask the candidate's permission to communicate with anyone else.

- Email and phone contact can be problematic. They communicate with their peers in different ways. Social media (?)
- Attention to detail may be limited. Our website which is often not seen, contains a great deal of information.
- Most candidates are very good and do everything the right way. A small minority do not.

Copyright 2025 Pearson. All rights reserved.

Approximately 17,000 candidates sit the test over a 5-week period. For most this is a straightforward process. For a relatively small number considerable guidance and support is required.

Copyright 2026 Pearson. All rights reserved.

Construct validation

Copyright 2026 Pearson. All rights reserved.

Validating what we measure

A UK-built test selects ANZ school-leavers, so validity evidence must be gathered locally.



One factor per cognitive subtest

Factor analysis every year supports reporting each subtest score separately.



Content globalised for ANZ

Cognitive items reviewed for language, context and units so they suit ANZ candidates.



SJT reviewed by ANZ experts

Scenarios rest on cultural and professional norms, so ANZ experts join development and concordance panels.



Evidence drives design

2025: AR removed (low predictive validity, high coachability); DM lengthened.

Copyright 2026 Pearson. All rights reserved.

Score comparability

Copyright 2026 Pearson. All rights reserved.

Building comparable forms

Five parallel forms each year (three in 2019); every candidate is randomly assigned one.

Results on the day: forms must be comparable before anyone sits the test.

	Cognitive (VR, DM, QR)	SJT
Statistically equivalent forms	Built from calibrated items	Built from items with known statistics
Content balance	Forms balanced on content as well as statistics	Forms balanced on content as well as statistics
Equating method	Rasch model: scores converted using existing UK item parameters	Classical pre-equating from item statistics
From 2026	Readability added as a form-balancing check (VR)	—

Copyright 2026 Pearson. All rights reserved.

Measurement quality

Copyright 2026 Pearson. All rights reserved.

Reliability: consistent across forms and years

Means across forms. Cognitive: scaled-score reliability. SJT: Cronbach's alpha.

*** 2019:** three forms.

† 2025: Abstract Reasoning removed, so the total is not directly comparable with earlier years.

Mean across forms	2019*	2020	2021	2022	2023	2024	2025†
VR	0.74	0.73	0.76	0.75	0.76	0.77	0.78
DM	0.71	0.75	0.77	0.78	0.78	0.78	0.80
QR	0.80	0.82	0.81	0.83	0.82	0.82	0.81
Total cognitive	0.91	0.92	0.93	0.93	0.93	0.94	0.92
SJT (alpha)	0.81	0.81	0.84	0.86	0.86	0.83	0.81

Copyright 2026 Pearson. All rights reserved.

Does a UK item bank fit an ANZ cohort?

Rule: items drifting beyond ±0.5 logits are recalibrated.

Few items: usually under 10% per subtest, apart from DM in 2020 and 2025. The 2020 DM spike was also seen in the UK.

No overall shift: mean difficulty within ±0.06 logits of the bank, except DM in 2020 (–0.12).



Copyright 2026 Pearson. All rights reserved.

Item-level fairness: few flags, some groups untested

Method:
Cognitive: Mantel-Haenszel D-DIF.

SJT: hierarchical regression, R-squared change.

What we found

- Cognitive (gender): one operational item a year rated C (2019–2024); none in 2025.
- SJT (gender, age): no items rated C in any year, 2019-2025
- Category C items are sent for content review

The gap

- Cognitive age DIF can't be computed: the under-20 vs over-35 grouping leaves too few older candidates

Copyright 2026 Pearson. All rights reserved.

From scores to selection policy

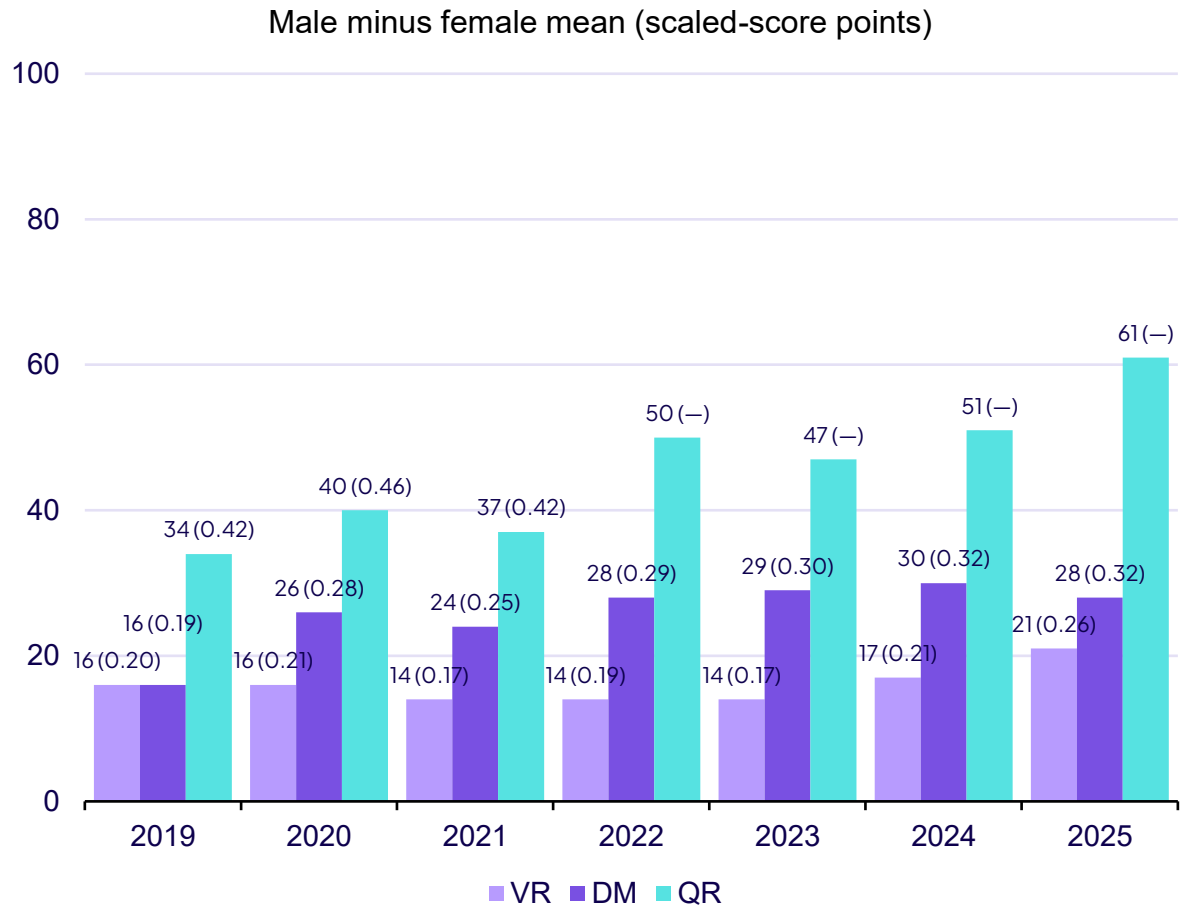
Copyright 2026 Pearson. All rights reserved.

Group differences: what they mean for selection

Consistent gaps: males score higher on the cognitive subtests; females score higher on the SJT (not shown).

SJT varies less across subgroups than the cognitive subtests.

Policy: universities should consider these gaps when deciding subtest weights and cut-offs.



Each subtest is scored 300–900. Cohen's d in brackets. — = not reported (group SDs differed significantly).

Copyright 2026 Pearson. All rights reserved.

Protecting what a score means



Exposure control

Five randomly assigned forms and embedded pretest items.



Drift monitoring

Item difficulty checked against the bank every year.



Coachability

AR removed partly because it was highly coachable.

Copyright 2026 Pearson. All rights reserved.

Comparable scores need valid constructs, comparable forms, ongoing evidence of reliability and fairness, and clear guidance on what group differences mean for selection.

Copyright 2026 Pearson. All rights reserved.

Pearson

Thank you. Questions?

edwardfeng.li@pearson.com

annette.mercer@monash.edu