MEKSI.AI

POST-EVENT REPORT

University of Melbourne OSCE Competition

Virtual-patient case: **Bob Ray** — a 55-year-old man presenting with breathlessness. Full consultation scored across Data Gathering, Data Interpretation and Clinical Decision Making.

Competition window
24 June – 4 July 2026 (AEST)

Platform
competition.meksi.com

Report generated
6 July 2026

Data source
Full production database export

Executive Summary

94	29	15	13
TOTAL SESSIONS	LEADERBOARD SUBMISSIONS	GENUINE ENTRANTS	FEEDBACK COMMENTS

35	1,913	3,815	93%
COMPLETED CONSULTATIONS	STUDENT QUESTIONS ASKED	TOTAL CHAT MESSAGES	WINNING SCORE (55/59)

Mitsuko Akiko won the competition with **55 / 59 (93%)** — the only entrant to top all three scored domains, including a near-perfect 7/8 in Data Interpretation and a perfect 11/11 in Clinical Decision Making. **Elizabeth Smith** (49, 83%) and **Michele Himadi** (47, 80%) complete the podium.

The competition drew **94 practice-and-attempt sessions** from 35 distinct devices over the event window, generating 3,815 chat messages. 29 scores were submitted to the leaderboard, of which **17 were genuine student entries from 15 entrants** (the remainder were internal test submissions, excluded from rankings and clearly flagged in Appendix A). Activity peaked on launch day (24 June, 25 sessions), with a second wave on 1 July and a final-day rush before the midnight 4 July deadline.

Domain performance at a glance

Entrants were strongest in Data Gathering (top scores 37/40) and Clinical Decision Making (a perfect 11/11 was achieved). Data Interpretation was decisively the hardest domain — the winner scored 7/8, but no other entrant exceeded 3/8. Student feedback independently corroborates this ("bit unclear on the data interpretation points"), making it the clearest curriculum signal of the event.

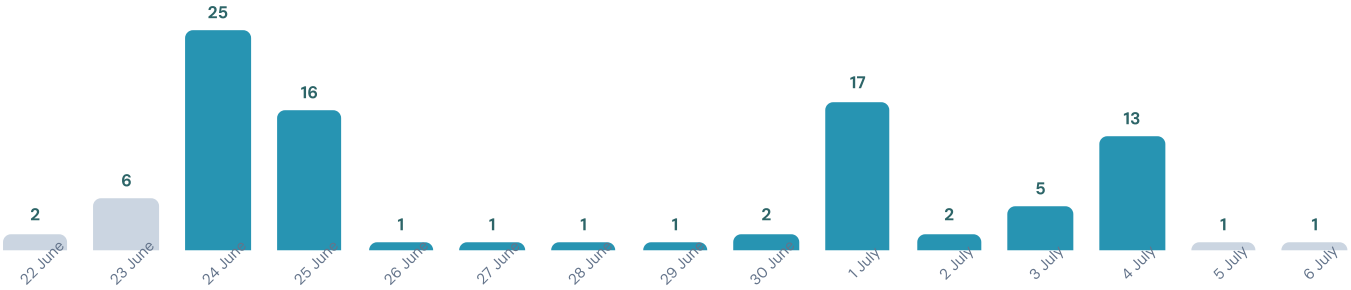
> **Scoring integrity note — two submissions were affected by a scoring-service outage.**
> On 1 July (18:00–19:10 AEST) the external intent-classification service exhausted its quota, so two end-of-session recomputes recorded 0 despite full consultations. The affected sessions' live scoring state was preserved and has been used to correct their entries in this report (flagged †): **Timothy Naylor 0 → 40** (would have ranked 5th) and Kelvin Ton's 1 July attempt 0 → 24 (superseded by his 4 July resubmission of 36 either way). Four further entries were undercounted by 1–2 points. Podium places 1–3 are unaffected. Details in Appendix B.

> **Student reception was positive.** Feedback highlights: *"Felt very realistic! Really helpful"*, *"I enjoyed doing this case"*, *"good case overall"*. Constructive themes: clarity of the Data Interpretation rubric, text-to-speech usability, occasional AI-patient confusion on single-symptom questions, and appetite for a second section / more cases.

Participation & Engagement

94	59	35	53
TOTAL SESSIONS	PRACTICE / UNFINISHED	COMPLETED	ENGAGED (≥3 QUESTIONS)

Sessions per day (AEST)



Grey bars fall outside the official window: 22–23 June = pre-launch smoke tests; 5–6 July = post-deadline stragglers. Competition window bars (24 June – 4 July) shown in MEKSI teal.

How to read the numbers

- **Sessions (94)** — every consultation started on the platform, including repeat practice runs. Students could restart freely; only a submitted attempt entered the leaderboard.
- **Submissions (29)** — scores pushed to the leaderboard: 17 genuine (15 distinct entrants; two entrants submitted twice) + 12 internal test rows.
- **Engagement** — 1,913 student questions across all sessions. Completed consultations ran a median of 33 questions; the most persistent entrant asked 265.

Scoring model

Each consultation is scored server-side against the SME-approved Bob Ray rubric: **59 points** across **Data Gathering (40)**, **Data Interpretation (8)** and **Clinical Decision Making (11)**. Every question a student asks is classified by an intent-recognition model and credited toward rubric tasks; a task is complete at half its intents (rounded up). From 24 June the case ran *free-order* scoring — questions score in any order, as in a real OSCE.

Final Leaderboard

Best submission per entrant, internal test rows removed, outage-affected entries corrected from preserved live scoring state (†). Tie-break: earlier submission ranks first.

	ENTRANT	SCORE	%	DATA GATHERING	DATA INTERP.	CLIN. DECISION	SUBMITTED (AEST)
🥇	Mitsuko Akiko	**55** / 59	93%	37/40	7/8	11/11	4 July 2026, 16:00
🥈	Elizabeth Smith	**49** / 59	83%	36/40	3/8	10/11	4 July 2026, 14:55
🥉	Michele Himadi	**47** / 59	80%	35/40	3/8	9/11	4 July 2026, 16:38 corrected †
4	Jackson	**41** / 59	69%	32/40	2/8	7/11	27 June 2026, 16:07
5	Timothy Naylor	**40** / 59	68%	32/40	2/8	6/11	1 July 2026, 18:03 corrected †
6	Samantha Braithwaite	**38** / 59	64%	31/40	2/8	5/11	30 June 2026, 16:14 corrected †
7	Kelvin Ton	**36** / 59	61%	33/40	2/8	1/11	4 July 2026, 16:49
8	Ann Guo	**35** / 59	59%	27/40	3/8	5/11	1 July 2026, 14:05
9	Bronwyn McDonald	**33** / 59	56%	26/40	2/8	5/11	3 July 2026, 12:29
10	Al	**28** / 59	47%	22/40	3/8	3/11	2 July 2026, 14:49
11	Rena Yeo	**27** / 59	46%	21/40	1/8	5/11	25 June 2026, 11:27
12	XAVIER	**26** / 59	44%	22/40	1/8	3/11	25 June 2026, 14:16
13	Lachlan Bowman	**24** / 59	41%	22/40	1/8	1/11	25 June 2026, 14:17 corrected †
14	Wanran Zhang	**24** / 59	41%	20/40	1/8	3/11	3 July 2026, 21:54
15	Farhan Islam	**21** / 59	36%	19/40	1/8	1/11	26 June 2026, 01:41 corrected †

† Corrected from the session's preserved live scoring state after the 1 July intent-service outage (see Appendix B). Podium unaffected. Eligibility screening of non-university email addresses is left to the organising team.

Category Leaderboards

Top 10 per scored domain — each entrant's best submission for that domain. † = outage-corrected entry.

Data Gathering / 40

	ENTRANT	SCORE	%
🥇	Mitsuko Akiko	**37**/40	93%
🥈	Elizabeth Smith	**36**/40	90%
🥉	Michele Himadi	**35**/40 †	88%
4	Kelvin Ton	**33**/40	83%
5	Jackson	**32**/40	80%
6	Timothy Naylor	**32**/40 †	80%
7	Samantha Braithwaite	**31**/40 †	78%
8	Ann Guo	**27**/40	68%
9	Bronwyn McDonald	**26**/40	65%
10	XAVIER	**22**/40	55%

Data Interpretation / 8

	ENTRANT	SCORE	%
🥇	Mitsuko Akiko	**7**/8	88%
🥈	Ann Guo	**3**/8	38%
🥉	Al	**3**/8	38%
4	Elizabeth Smith	**3**/8	38%
5	Michele Himadi	**3**/8 †	38%
6	Jackson	**2**/8	25%
7	Samantha Braithwaite	**2**/8 †	25%
8	Timothy Naylor	**2**/8 †	25%
9	Kelvin Ton	**2**/8 †	25%
10	Bronwyn McDonald	**2**/8	25%

Clinical Decision Making / 11

	ENTRANT	SCORE	%
🥇	Mitsuko Akiko	**11**/11	100%
🥈	Elizabeth Smith	**10**/11	91%
🥉	Michele Himadi	**9**/11 †	82%
4	Jackson	**7**/11	64%
5	Ann	**6**/11	55%
6	Timothy Naylor	**6**/11 †	55%
7	Rena Yeo	**5**/11	45%
8	Samantha Braithwaite	**5**/11 †	45%
9	Bronwyn McDonald	**5**/11	45%
10	XAVIER	**3**/11	27%

Observations

- **Data Gathering** — the strongest domain across the field: seven entrants scored 20+/40 and the top three were within 2 points of each other. History-taking skills translated well to the AI-patient format.
- **Data Interpretation** — the discriminator. Mitsuko's 7/8 stands alone; every other entrant scored ≤3/8. Students reported being unsure how to voice interpretation to the patient — a rubric-communication gap worth addressing before the next event.
- **Clinical Decision Making** — a perfect 11/11 was achieved (Mitsuko), with a clear podium gradient below; mid-field entrants often ran out of time before management.

Student Feedback

All feedback received via the in-app feedback form, deduplicated (the form allowed rapid resubmission — duplicates of the identical comment are collapsed with a ×N marker). 13 unique student comments; internal test submissions listed separately below.

FIRST SUBMITTED	FROM	FEEDBACK
25 June 2026, 14:17	Lachlan Bowman	ABC submitted ×3
30 June 2026, 16:11	Samantha Braithwaite	i liked the case i found it good however i found the text to speech hard to use so i had to use a notes page so i could speak outloud. unfortunatly as seen in my transcript I tried using it once and it sent the start of my sentence twice. I also would like to say how cool of a System this is I would potentially in the future seek a little bit more clarification on how to get more marks in clinical decision and data interpretation as I wouldn't explain to the patient why? For example the ECG came back suggesting a previous interior am I. Which would explain some of his heart symptoms through ischemic cardiomyopathy. maybe this could be a second section? submitted ×2
30 June 2026, 16:12	Samantha Braithwaite	i liked the case i found it good however i found the text to speech hard to use so i had to use a notes page so i could speak outloud. unfortunatly as seen in my transcript I tried using it once and it sent the start of my sentence twice. I also would like to say how cool of a System this is I would potentially in the future seek a little bit more clarification on how to get more marks in clinical decision and data interpretation as I wouldn't explain to the patient why? For example the ECG came back suggesting a previous interior am I. Which would explain some of his heart symptoms through ischemic cardiomyopathy. maybe this could be a second section?
30 June 2026, 16:14	Samantha Braithwaite	I enjoyed doing this case, however I would like potentially second section just so that I could explain my interpretation as I wouldn't explain this to the patient. And I found the text to speech function unusable unfortunately.
1 July 2026, 18:00	Timothy Naylor	Felt very realistic! Really helpful submitted ×3
1 July 2026, 18:40	Kelvin Ton	Good case. The score stopped increasing even though I feel I was asking appropriate questions? Not too sure submitted ×2
3 July 2026, 21:51	Wanran Zhang	-
4 July 2026, 14:55	Elizabeth Smith	bit unclear on the data interpretation points - hard to give patient interpretation of data without all information submitted ×2
4 July 2026, 14:57	Elizabeth Smith	bit unclear on the data interpretation points - hard to give patient interpretation of data without all information submitted ×2
4 July 2026, 15:52	Mitsuko	I broke the AI a few times by just asking about single symptoms. submitted ×3
4 July 2026, 15:54	Mitsuko Akiko	I broke the AI a few times.
4 July 2026, 16:36	Michele Himadi	Nothing so far, but I attempted this earlier and thought the rubric may seem a bit inconsistent. submitted ×2
4 July 2026, 16:48	Kelvin Ton	I think good case overall, but quite unclear of where to go at times submitted ×3

Internal test rows (excluded)

FIRST SUBMITTED	FROM	FEEDBACK
24 June 2026, 07:40	A	Good
24 June 2026, 08:45	B	SOme good

FIRST SUBMITTED	FROM	FEEDBACK
25 June 2026, 15:40	Bishal	Bishal
3 July 2026, 20:46	bishal	This was good
3 July 2026, 22:43	test	dasf

Appendix A — Raw Leaderboard (as recorded)

Every leaderboard row as stored in the production database, in recorded score order, including internal test submissions (greyed) and pre-correction scores. "Live" = the score the student saw at the end of their session (preserved server-side scoring state), shown where it differs meaningfully from the recorded recompute.

SUBMITTED (AEST)	NAME	RECORDED	DG	DI	CDM	LIVE	FLAGS
4 July 2026, 16:00	Mitsuko Akiko	55 / 59	37	7	11	55	
4 July 2026, 14:55	Elizabeth Smith	49 / 59	36	3	10	48	
4 July 2026, 16:38	Michele Himadi	46 / 59	35	3	8	47	+1 undercount
27 June 2026, 16:07	Jackson	41 / 59	32	2	7	40	
30 June 2026, 16:14	Samantha Braithwaite	37 / 59	30	2	5	38	+1 undercount
4 July 2026, 16:49	Kelvin Ton	36 / 59	33	2	1	35	
1 July 2026, 14:05	Ann Guo	35 / 59	27	3	5	35	
3 July 2026, 12:29	Bronwyn McDonald	33 / 59	26	2	5	33	
30 June 2026, 17:34	Ann	32 / 59	24	2	6	32	
2 July 2026, 14:49	Al	28 / 59	22	3	3	28	
25 June 2026, 11:27	Rena Yeo	27 / 59	21	1	5	27	
25 June 2026, 14:16	XAVIER	26 / 59	22	1	3	26	
3 July 2026, 21:54	Wanran Zhang	24 / 59	20	1	3	24	
25 June 2026, 14:17	Lachlan Bowman	23 / 59	22	1	0	24	+1 undercount
26 June 2026, 01:41	Farhan Islam	19 / 59	17	1	1	21	+2 undercount
24 June 2026, 07:40	A	3 / 58	3	0	0	3	test
25 June 2026, 15:40	Bishal	3 / 59	2	1	0	3	test
23 June 2026, 14:04	Bishal Sapkota	2 / 58	2	0	0	2	test
24 June 2026, 08:45	B	2 / 58	2	0	0	2	test
3 July 2026, 20:46	bishal	2 / 59	2	0	0	2	test
24 June 2026, 13:03	Amaan	1 / 58	1	0	0	1	test
3 July 2026, 22:43	test	1 / 59	1	0	0	1	test

SUBMITTED (AEST)	NAME	RECORDED	DG	DI	CDM	LIVE	FLAGS
4 July 2026, 21:58	s	1 / 59	0	0	1	1	test
1 July 2026, 18:03	Timothy Naylor	0 / 59	0	0	0	40	+40 undercount
1 July 2026, 19:05	Kelvin Ton	0 / 59	0	0	0	24	+24 undercount
2 July 2026, 12:42	Bishal	0 / 59	0	0	0	0	test
4 July 2026, 16:42	tes	0 / 59	0	0	0	—	test
4 July 2026, 21:56	a	0 / 59	0	0	0	0	test
4 July 2026, 21:57	a	0 / 59	0	0	0	0	test

Appendix B — Scoring Integrity & Corrections

What happened

Leaderboard scores are produced by an end-of-session *recompute*: the full transcript is re-scored from scratch through the external intent-classification service (an anti-tamper measure — the ranked score is derived only from the saved transcript). Between 1–2 July that external service exhausted its API quota; recomputes during the outage silently credited nothing. The **live** scoring state each student accumulated during their consultation was unaffected and remained in the database, which is what allows a faithful correction.

Affected submissions (live state > recorded score)

ENTRANT	SUBMITTED (AEST)	RECORDED	LIVE STATE	Δ
Michele Himadi	4 July 2026, 16:38	46	**47**	+1
Samantha Braithwaite	30 June 2026, 16:14	37	**38**	+1
Lachlan Bowman	25 June 2026, 14:17	23	**24**	+1
Farhan Islam	26 June 2026, 01:41	19	**21**	+2
Timothy Naylor	1 July 2026, 18:03	0	**40**	+40
Kelvin Ton	1 July 2026, 19:05	0	**24**	+24

- **Timothy Naylor (0 → 40)** is the material correction: a full consultation scored during the outage. Corrected, he ranks **5th**. His feedback that evening — "Felt very realistic! Really helpful" — was submitted minutes before the failed recompute.
- **Kelvin Ton's 1 July attempt (0 → 24)** was also outage-hit, but his 4 July resubmission (36) is his best entry regardless.
- The four ±1–2 point deltas are minor classification variance between live and recompute passes, listed for completeness. No ranking changes at the podium.

Other data notes

- **Scoring mode change mid-event (24 June):** the case was promoted from sequential to free-order scoring on launch morning with SME sign-off; all genuine leaderboard entries were scored under free-order rules (rubric v5), so rankings are like-for-like.
- **Test data:** 12 leaderboard rows and 5 feedback rows from internal test emails are flagged and excluded from all rankings.
- **Duplicate entrants:** two entrants submitted twice; rankings use each entrant's best submission.
- **Deadline:** last genuine submission 4 July 16:49 AEST, within the midnight deadline. (Three test rows landed 21:56–21:58 that evening.)
- **Unsubmitted efforts:** a handful of substantial consultations were never submitted to the leaderboard — the largest reached 29/59 after 48 questions. Nothing rank-relevant.