Assessing the Educational Value of an AI Tutor in FRCR 2B Short Case Preparation Using Longitudinal Performance Analytics

Girija Agarwal[1], Naila Hamrioui[2,3], Paymon Zomorodian[2,4]

[1] Imperial College Healthcare NHS Trust [2] RadBytes [3] Warrington and Halton Teaching Hospitals NHS Foundation Trust [4] Mersey and West Lancashire Teaching Hospitals NHS Trust

Key message

Cubey use was associated with a statistically significant, modest improvement in FRCR 2B short case scores over time. Learners who improved more also interacted with the AI tutor more frequently per case.

7,944
scored cases analysed

152
users included

β = 0.000301
score increase per case; p = 0.013

3.41 vs 2.96
interactions per case in higher- vs low-improvement users; p < 0.05

93.8%
survey respondents felt Cubey improved performance

BACKGROUND

FRCR 2B short cases preparation relies on repeated practice, feedback and calibration against exam-style marking criteria for radiograph interpretation. However, access to timely senior feedback can be variable.

AI-supported tutoring may offer a scalable way to provide structured feedback, helping learners identify weaknesses, refine reporting style, and benchmark progress against exam-style criteria.

Cubey is an AI-supported radiology tutor within the RadBytes platform. It provides structured feedback on anonymised educational short cases and assigns scores using standardised FRCR 2B marking criteria.

AIM

To evaluate whether interaction with the Cubey AI tutor is associated with measurable improvement in user performance for short case radiograph reporting and whether engagement behaviour differs between higher- and lower-improvement learners.

HOW CUBEY WAS USED

The platform contained exam-style short cases, which includes a radiograph and short clinical history. Learners reviewed each case and submitted a report, following which they received an AI-generated score and feedback, and could continue iterating through further messages with the AI tutor.

METHODS

A longitudinal retrospective analysis was conducted using anonymised interaction data from users of the Cubey FRCR 2B short case platform.

Only cases receiving a formal AI-assigned score (1–5) were included.

Performance was analysed using a linear mixed-effects model with score as the dependent variable, case order as a fixed effect, and user as a random intercept.

Individual learning slopes were also calculated to classify users as higher-improvement or low-/non-improvement.

The number of messages exchanged between the learner and Cubey, i.e. the number of "interactions" was treated as a marker of engagement. Total interactions per case were compared between these groups using a t-test to assess differences in engagement behaviour.

RESULTS: COHORT AND ENGAGEMENT

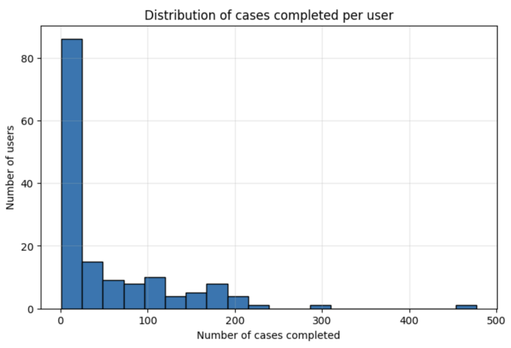
A total of 152 users submitted 7,944 scored cases. Engagement varied substantially, reflecting real-world use from brief trialling to sustained exam preparation.

A median of 17.5 cases were done per user (IQR 3.0–88.25) and a maximum of 477 cases were done by the highest user.

Median cases done per user

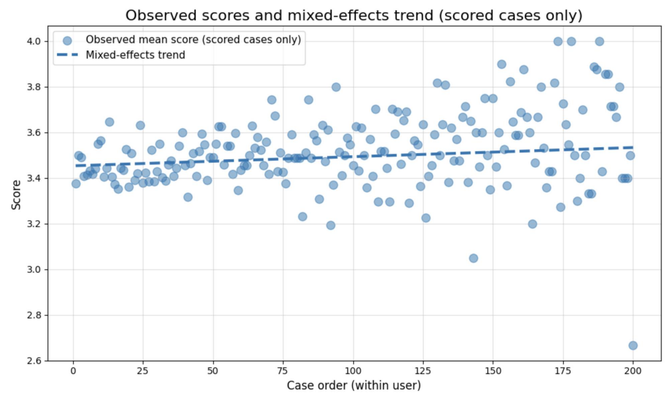
17.5

IQR 3–88.25 | maximum 477 cases



RESULTS: PERFORMANCE TREND

Mixed-effects modelling showed a statistically significant positive association between case order and score (β = 0.000301 per case, p = 0.013), indicating gradual improvement with increased case exposure.

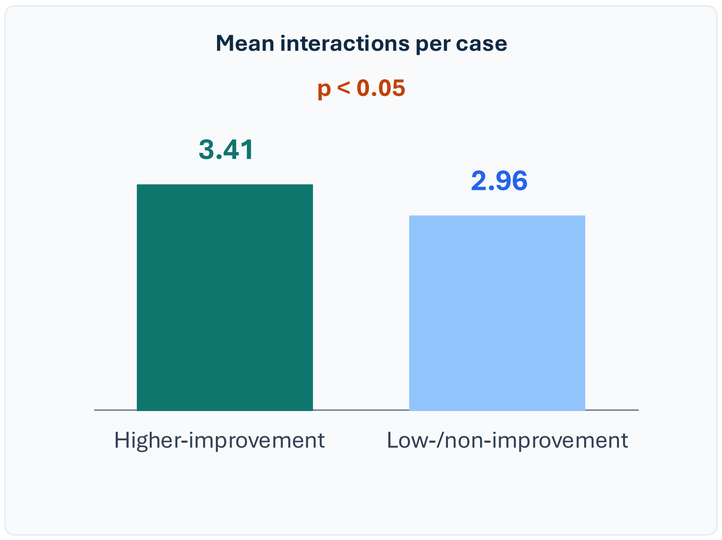


In this plot, the x-axis represents the user's nth completed case (case order), while the y-axis represents the mean score across all users for that corresponding case order, thereby illustrating how average performance changes as users complete increasing numbers of cases.

Cases beyond a case order of 200 were excluded from graphical analysis due to sparse data at higher case numbers.

RESULTS: ENGAGEMENT BEHAVIOUR

Higher-improvement users demonstrated significantly greater interaction volume compared to low/non-improvement users, with a mean of 3.41 vs 2.96 interactions per case respectively ($p < 0.05$) suggesting that more engagement with the AI tutor was associated with improved performance.



DISCUSSION

The score change was statistically significant but modest in absolute size, which is realistic for exam preparation data.

Engagement appears important: users who interacted more per case showed greater improvement.

The findings support AI feedback as repeated formative practice, not as a replacement for consultant supervision or peer viva practice.

Practical takeaway

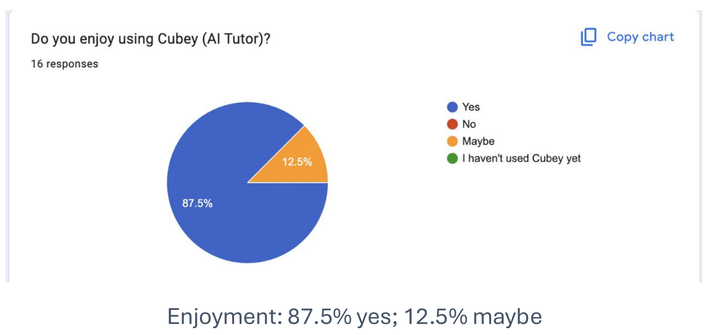
The educational value appears strongest when learners actively iterate with the tutor rather than passively submitting isolated cases.

USER FEEDBACK

Post-use survey responses were strongly positive, supporting learner acceptability alongside the quantitative described.

Of 16 survey responses, 93.8% of respondents felt that Cubey improved their short case reporting and performance, and 87.5% reported enjoying using it.

Perceived improvement: 93.8% yes



Enjoyment: 87.5% yes; 12.5% maybe

EDUCATIONAL IMPLICATIONS

AI tutors may help learners practise more cases between formal teaching sessions.

Structured feedback may help candidates recognise recurring reporting gaps and calibrate to marking criteria.

Engagement metrics could identify learners who may benefit from nudges, targeted support or additional human teaching.

CONCLUSION

Use of an AI tutor in FRCR 2B short case preparation is associated with statistically significant though modest improvement in performance over time.

Notably, users who demonstrated greater improvement engaged more frequently per case with the AI tutor.

In addition, users feedback regarding the utility of the tool was positive. These findings support the role of AI-driven tutoring platforms as effective adjuncts to radiology exam preparation.

LIMITATIONS AND FUTURE WORK

Retrospective platform analysis; causality cannot be inferred.

Motivated users may both submit more cases and improve more, creating selection bias.

Future work should compare AI scores with human examiner scores and evaluate outcomes prospectively.

DISCLOSURES

NH and PZ are co-founders of RadBytes

.