

Deep Reinforcement Learning for Engagement-Aware Question Selection in Adaptive Assessment Systems

Okan Bulut
University of Alberta

Joyce Xinle Liu
University of Alberta

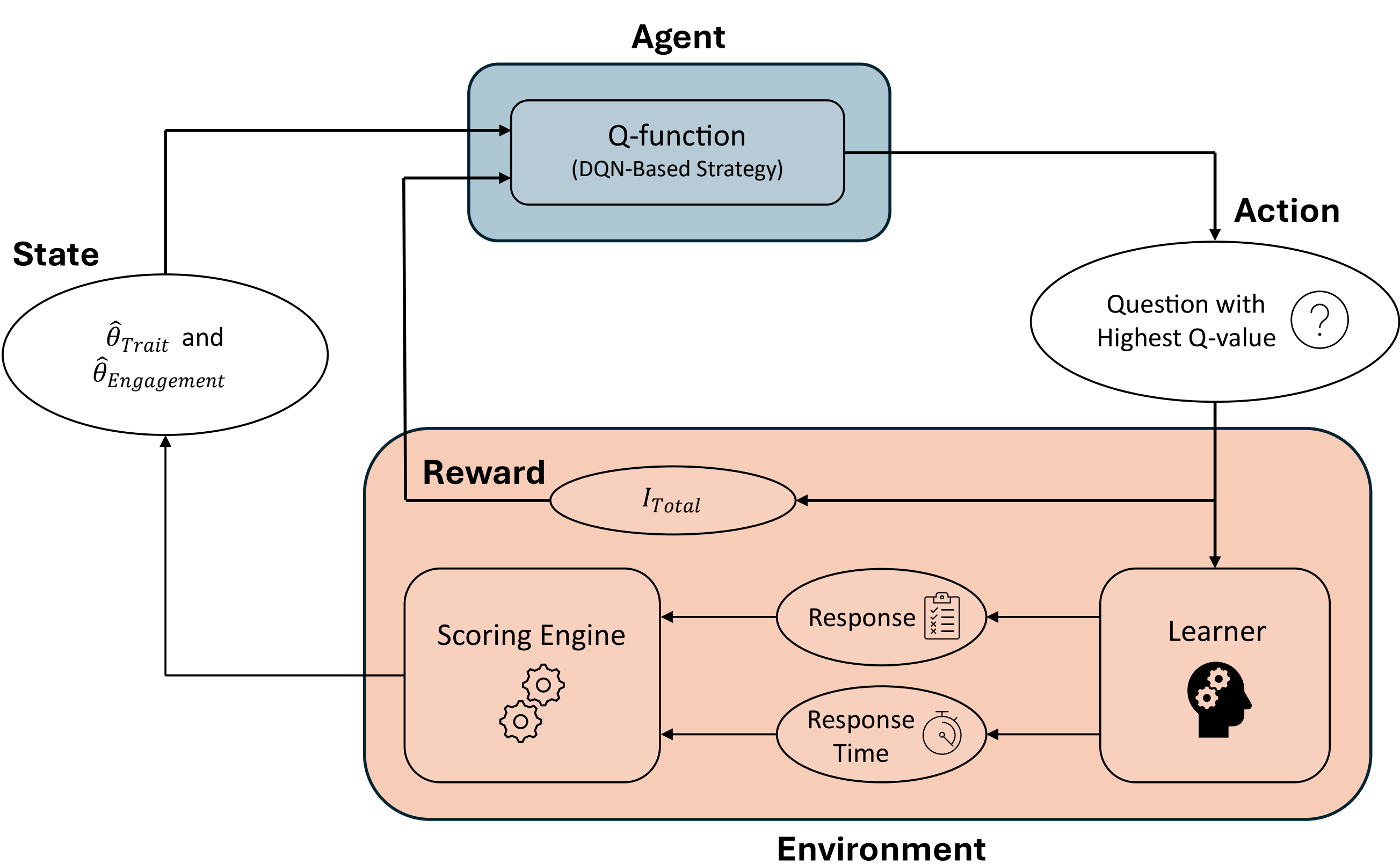
Hümeyra Demir
Hacettepe University

INTRODUCTION

- **Adaptive assessments** dynamically select questions based on a learner's responses and estimated ability level, offering improved efficiency and personalization compared to fixed-form tests.
- However, **disengaged behaviors** (e.g., rapid guessing, idling) can compromise the accuracy of ability estimates in **item response theory (IRT)** models and disrupt the question selection process.
- Leveraging **reinforcement learning (RL)**, this study proposes a **proactive approach** to monitor the learner's engagement level in **real time** to guide question selection.

METHODS

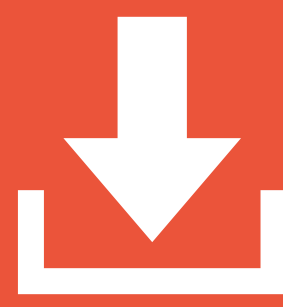
- **Empirical dataset:** $n = 22,653$ Canadian students (Grades 8, 9). 244 multiple-choice questions, each with a binary response and **response time**.
- **Engagement modeling:** Engagement was modeled as a latent trait (Rasch model) based on three indicators identified using question-level time thresholds: 0=rapid guessing or idling vs. 1=effortful responses
- **Performance modeling:** 2PL IRT model
- **Question selection:** **Deep Q-Network (DQN)** approach that selects questions based on performance and engagement to maximize **long-term information gain**.
- **Post-Hoc Simulation:** DQN vs. IRT-based selection



DISCUSSION

- By incorporating engagement into the question selection process, the DQN-based method can improve the **accuracy** and **efficiency** of adaptive assessments.

Reinforcement Learning can balance learner performance and engagement to ensure timely completion in adaptive assessments.



Scan the QR code to download this poster.

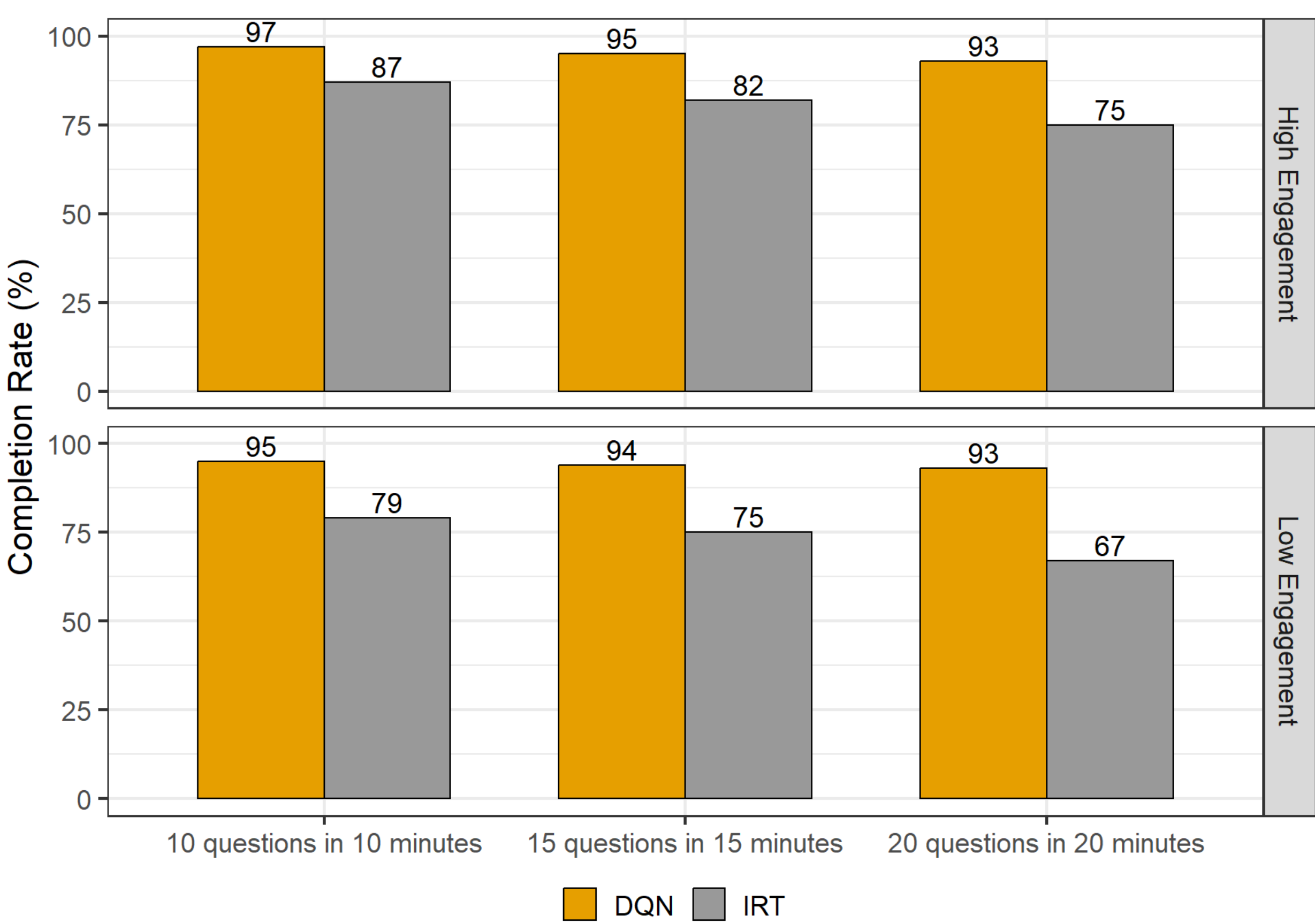


RESULTS

- **DQN outperformed IRT** in all three scenarios: lower RMSE and higher correlation with true trait levels.

Scenario	Method	Bias	RMSE	Corr.
10 questions within 10 minutes	DQN	0.017	0.448	0.951
	IRT	0.021	0.587	0.873
15 questions within 15 minutes	DQN	0.009	0.366	0.962
	IRT	0.014	0.451	0.917
20 questions within 20 minutes	DQN	0.003	0.291	0.977
	IRT	0.005	0.388	0.939

- **Test Completion Rates:** $\geq 93\%$ completion rate for DQN in all scenarios, consistently higher than that for the IRT-based method. As the number of questions increased, the completion rate under DQN remained stable, while the IRT-based method saw substantial declines.



- **Future work:** Extension to multidimensional adaptive assessments, incorporation of additional process data to obtain a more nuanced and robust understanding of engagement.



Dr. Okan Bulut
bulut@ualberta.ca



UNIVERSITY
OF ALBERTA



HACETTEPE
UNIVERSITY