

Feedback 1: Clarification on the Use of Student Embeddings

- **Verbatim Feedback Received:** "i'm not sure what to make of user embeddings. i guess i can't tell if you have gone from predicting difficulties *WITHOUT* looking at person responses to something more mixed?"
- **Evaluation of Feedback:** This feedback was valuable as the role of student embeddings is crucial for understanding how the model simulates student responses for unseen questions.
- **Description of Revisions:** The section on Student Embeddings (Section 4.1.5) was revised to more clearly articulate how these learned embeddings are utilized. The text now emphasizes that once the student ability profiles are learned from the training data, the model can leverage these embeddings to predict how the cohort of known students would likely respond to any new, unseen question. This process generates a matrix of "synthetic student responses," which is fundamental to the subsequent IRT analysis for estimating difficulty without requiring new real-world student testing. This contrasts with the previous draft's explanation, which focused more on the technical implementation and the benefits over one-hot encoding, rather than its direct application in the simulation stage for new items.
- **Justification of Revisions:** The clarification ensures that readers understand the mechanism by which student-specific characteristics are carried forward to estimate difficulty for new items, reinforcing the paper's core methodology of simulating student responses.
- **Reference in Manuscript:** The updated explanation can be found in Section 4.1.5, "Student Embeddings".

Feedback 2: Introduction of an RMSE-Based Efficiency Metric

- **Paraphrased Feedback Received:** It was suggested that incorporating an efficiency metric, specifically one based on Root Mean Square Error (RMSE), would be beneficial. This metric would help quantify the practical value of the model by comparing its difficulty estimation accuracy to traditional IRT methods that rely on varying amounts of actual student data.
- **Evaluation of Feedback:** This was an excellent suggestion, as it provides a concrete way to demonstrate the model's utility and potential for resource savings.
- **Description of Revisions:** A new subsection titled "RMSE-Based Efficiency Evaluation" (Sec-

tion 5.3) was introduced. This section details the methodology where the RMSE of our model's difficulty predictions (on holdout questions, using no new student data for these specific items) is compared against the RMSE achieved by a standard 1PL IRT model using progressively larger subsamples of actual student responses for those same holdout questions. A new figure (Figure 7) was created to visually represent this comparison, showing the number of real student answers that would be needed by a traditional IRT approach to achieve the same RMSE as our model.

- **Justification of Revisions:** This addition significantly strengthens the paper by providing a quantitative measure of the model's efficiency. It demonstrates the practical implications of the research, showing how the model can achieve robust difficulty estimates with potentially fewer real student interactions than traditional pre-testing.
- **Reference in Manuscript:** This new analysis is presented in Section 5.3, "RMSE-Based Efficiency Evaluation" and visually supported by Figure 7.

Feedback 3: Minor Wording and Terminological Adjustments

- **Paraphrased Feedback Received:** Recommendations to enhance clarity by preferring "problem statement" over "question title" in specific contexts and introducing "IRT difficulty" earlier in the paper.
- **Evaluation of Feedback:** These suggestions improved the manuscript's readability and precision.
- **Description of Revisions:** Terminology was updated throughout for consistency. "Question title" was changed to "problem statement" where appropriate (e.g., in Section 4.1.1), and "IRT difficulty" was explicitly defined earlier in the Abstract and Introduction.
- **Justification of Revisions:** These changes ensure consistent terminology and appropriate introduction of key concepts.
- **Reference in Manuscript:** Changes appear in Section 4.1.1 "Question Text Embeddings", the Abstract, and Section 1 (Introduction).