# Bridging Human Insight and AI in Education: Teacher Observations for Academic Performance Prediction and Retrieval-Augmented Assessment

メンナトアッラヒ ファティーン

（様式 2）

氏　　名　：MENATALLAH FATEEN
メンナトアッラヒ ファティーン

論 文 名　：Bridging Human Insight and AI in Education: Teacher Observations for Academic Performance Prediction and Retrieval-Augmented Assessment
(人間の洞察力と AI を繋ぐ：学習能力予測への教師観察と検索強化型評価を例として)

区　　分　：　甲

## 論 文 内 容 の 要 旨

Educational assessment often reduces student performance to numerical scores. While these scores are useful for benchmarking, they fail to capture the deeper context of student learning observed by educators or provide the detailed, formative feedback essential for learning. Similarly, AI tools that aim to predict student performance or automatically assess answers often focus on quantitative metrics only. This thesis explores how integrating human expertise with AI-based tools can address this gap, focusing on two core areas: student performance prediction and automatic short-answer scoring with feedback.

The first contribution addresses the challenge of predicting student performance by incorporating qualitative teacher observation reports, which often contain rich contextual insights that are typically overlooked by traditional models. We utilize a unique dataset that combines student performance data with teacher observation reports, which provide detailed qualitative feedback on students' academic progress. Using novel representation learning techniques, including sentiment-based embeddings with similarity learning, we develop the Teacher Observation for Performance Prediction (TOPP) framework which effectively integrates these qualitative teacher insights with quantitative scores to enhance predictive accuracy. Specifically, we introduce four main representation methods that use the teacher observation reports to predict student performance in distinct ways. For each model, we compare the performance with and without the teacher observations, utilizing three main feature sets. Evaluated across the four modeling approaches, TOPP consistently improved prediction accuracy—achieving up to a 32% increase in performance—demonstrating how effectively human expertise can enhance AI-driven educational assessment.

The second contribution advances automatic short answer scoring (ASAS) through two complementary studies. The first study compares instance-based and reference-based scoring methods, examining their effectiveness in both traditional and zero-shot learning scenarios. While instance-based methods achieve higher accuracy in traditional scenarios, reference-based approaches maintain more consistent performance across both contexts—a

crucial insight for deploying ASAS systems in data-scarce environments.

Building on these insights, we introduce a novel retrieval-augmented generation framework that dynamically incorporates human-scored examples to provide both scores and detailed feedback (ASAS-F-RAG). By aligning with human judgment while minimizing fine-tuning requirements, our approach outperforms traditional baselines—achieving 9% higher accuracy on unseen questions and 1% on unseen answers—while reducing computational demands.

This thesis establishes new methodologies for meaningfully integrating human expertise into AI-powered educational assessment. Through teacher observations and retrieval-augmented techniques, we demonstrate how AI systems can better capture educational context, enhance prediction accuracy, and provide detailed feedback—advancing toward more responsive learning environments that benefit both educators and students.