AAC_S25 - Joe Carrier - 8675309 - Project Proposal

Comprehensive AI-Assisted Writing Assessment Using Ten Band Rubric and a Taxonomy of Errors

Scope: This study investigates the implementation of an AI-assisted assessment system using a ten-band analytic rubric for evaluating discourse competencies and a taxonomy of errors for evaluating mechanical and grammatical accuracy in argumentative essays produced by Korean EFL university students. A digital assessment system (DAS) coded in Python with multiple API calls to selected LLMs will conduct detailed analysis using the taxonomy and rubric. The analysis will focus on quantifiable linguistic features, including spelling errors and grammatical inaccuracies (e.g., subject-verb agreement, tense shifting, article misuse). Data collection will involve three longitudinal writing samples from a single cohort, analyzed through natural language processing pipelines trained on error pattern recognition. Lexical diversity and stylistic sophistication fall outside the study's parameters, as does the evaluation of non-textual elements like formatting or visual presentation.

Importance: Existing writing assessment rubrics often conflate mechanical accuracy with broader communicative competence, creating reliability challenges in EFL contexts where foundational language skills require targeted intervention. This study addresses a critical gap in second language acquisition research by developing a granular, AI-compatible framework that isolates developmental patterns in grammatical and orthographic proficiency. The ten-band scale enables precise tracking of incremental progress often obscured by traditional four-category systems, particularly valuable for learners exhibiting fossilized error patterns. Practical applications include automated feedback systems for large writing courses and diagnostic tools identifying persistent L1 interference issues in Korean-English language transfer. Educational institutions may adopt this model to standardize mechanical accuracy assessments while reducing grader workload through AI-assisted error tagging. The taxonomy's machine-readable structure further supports adaptive learning systems targeting individual error frequency profiles.

Methodology: Longitudinal data will be collected through timed writing tasks administered during weeks 1, 8, and 15 of a 15-week academic semester under examination conditions (no AI tools allowed). Participants will produce short paragraphs under uniform prompt conditions. A bidirectional LSTM model pretrained on 5,000 human-graded EFL essays will perform initial error calibration. Recursive improvements will be made in the taxonomy at this stage. The student writings will then be assessed by the DAS using the rubric and taxonomy, followed by manual validation using a subset of 27 student writings (9 per collection phase). Inter-rater reliability will be established through parallel scoring by three certified ESL assessors, with Cohen's kappa coefficients calculated for each rubric band. Error frequency counts will be normalized per 100 words to account for text length variations, while syntactic complexity metrics will be derived from dependency parse trees.

Literature Review: The literature review will analyze three primary domains: historical developments in ESL rubric design, computational approaches to error detection, and psycholinguistic models of L1 interference. Foundational works by Jacobs et al. on compositional assessment and Hyland's research into multi-band analytic rubrics will establish theoretical frameworks for the ten-band structure. Contemporary studies on neural network applications in grammatical error correction will be evaluated, particularly those addressing agglutinative language transfer issues. A critical analysis of rubric optimization research will compare holistic

versus analytic scoring systems, with emphasis on Sadler's threshold concepts for criterion-referenced assessment. The review will also synthesize findings from Korean EFL error analysis studies, particularly those documenting fossilization patterns in article usage and tense consistency. Emerging methodologies in automated writing evaluation will be examined through meta-analyses of validation studies employing quadratic weighted kappa statistics. This structured approach ensures comprehensive coverage of rubric design theory, AI-assisted assessment technical literature, and domain-specific research on Korean learner English. The synthesis will highlight underexplored intersections between granular banding systems and machine learning architectures optimized for morphological analysis.