**Enhancing Rubric-Based Evaluations in Educational Settings through Artificial Intelligence: An Extension of the Knowledge-Learning-Instruction (KLI) Framework**

Marcus D. Taylor

Learning Technologies Department, University of North Texas

LTEC 6000: Philosophy of Computing in Learning Technologies

Dr. Cathleen Norris

October 16, 2023

**Introduction**

Rubric-based assessments are widely used in educational settings to evaluate student performance against established criteria and standards (Reddy & Andrade, 2010). Rubrics are also useful to help students develop a critical sense of their own work by providing them with criteria to become more thoughtful judges of the quality of their own and others' work. True assessment emphasizes the application and use of knowledge to solve complex tasks that involve contextualized problems. Rubrics help students to understand the criteria for judgment from the beginning of their instruction (Montgomery, 2002). However, concerns around the reliability and validity of human-scored rubric evaluations persist due to issues of subjectivity and inconsistency (Panadero & Jonsson, 2013). Recent advancements in Artificial Intelligence (AI) technologies present new opportunities to address these limitations through automation and enhanced objectivity.

This research explores the integration of AI, specifically machine learning and natural language processing, into rubric-based assessments. It aims to investigate: (1) how AI can improve reliability and validity in evaluations; (2) AI's capability to provide personalized feedback and scaffolding via frameworks like Knowledge-Learning-Instruction (KLI); and (3) the ethical considerations and challenges that must be navigated. A mixed methods approach combines literature analysis, quantitative comparisons of AI and human scoring, and qualitative studies of real-world implementation.

The study fills a gap in examining how AI might transform rubric-based evaluations by extending the KLI framework – which has yet to incorporate AI elements. Findings promise to inform the balanced and ethical integration of AI in enhancing educational assessments. Practical implications span multiple stakeholders including educators, administrators, and policy makers exploring this rapidly evolving technological space. An integrated system utilizing AI to create, evaluate, and enhance rubrics for personalized and efficient educational processes (The Knowledge-Learning-Instruction (KLI) framework emerges as a comprehensive model for effectively integrating AI technologies into rubric-based evaluations. By extending the KLI framework, dynamic and tailored assessments can be realized. This not only promises to improve the reliability of the assessments but also to provide more personalized feedback and scaffolding to learners, thereby enhancing the overall educational experience (Nkambou et al., 2010).

However, while the potential benefits of integrating AI into educational evaluations are manifold, this approach also introduces new complexities and challenges. These include but are not limited to, the need to maintain the validity of assessments and ethical considerations such as data privacy and potential algorithmic bias. Therefore, for the successful deployment of AI in educational evaluations, a balanced and thoughtful approach is crucial.

**Literature Review**

**The Problem**

The integration of artificial intelligence (AI) technologies into educational assessments and evaluations introduces both opportunities and challenges agreement (Panadero & Jonsson, 2013; Reddy & Andrade, 2010). While AI-enabled evaluations promise to improve reliability, objectivity, and personalization (Nkambou et al., 2010), they also raise complex issues around maintaining validity, data privacy, and avoiding algorithmic bias. As educators look to leverage the benefits of AI, thoughtful frameworks like the Knowledge-Learning-Instruction (KLI) model are needed to ensure AI is deployed ethically and effectively. However, research needs comprehensive guidelines for integrating AI in a manner that maximizes benefits while proactively addressing risks. This presents a critical problem, as educators require evidence-based guidance on if and how to utilize AI technologies within assessments in a balanced, equitable, and pedagogically sound manner. The proposed literature review will synthesize current research on the use of AI in educational evaluations to identify best practices, limitations, and open questions. This synthesis will highlight key considerations for stakeholders exploring the integration of AI and provide recommendations for further research to promote responsible AI deployment in the education sector.

**Reliability and Validity of Rubric Evaluations**

A persistent issue with rubric scoring is reliability - consistency across raters and occasions (Jonsson & Svingby, 2007). Estimates of interrater reliability for rubrics range considerably from moderate agreement to almost perfect agreement (Panadero & Jonsson, 2013; Reddy & Andrade, 2010). Sources of variance include interpretation differences, scoring biases, and drift over time.

AI techniques like machine learning offer opportunities to improve reliability. AI algorithms can be trained on datasets of human-scored evaluations to build more consistent scoring models Researchers have developed automated essay-scoring systems that approach or surpass human levels of reliability on writing assessments (Shermis, 2014). AI techniques like clustering, classification, and natural language processing can evaluate textual elements more objectively.

However, increased reliability does not necessarily confer greater validity in measuring skills and knowledge (Williamson et al., 2012).AI models may reproduce human biases, limiting validity. Multidimensional rubrics provide more diagnostic information on constructs of interest than single scores (Jonsson & Svingby, 2007). AI techniques should be applied carefully to provide valid insights into the targeted knowledge, skills, and competencies.

Additional reliability issues with rubrics stem from individual scoring biases and drift over time, affecting consistency and objectivity (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). AI scoring offers more objective evaluation models trained on human-scored rubric data (Shermis, 2014). However, improved reliability does not always improve validity in skills measurement (Williamson et al., 2012). A limitation of AI scoring is the ability to provide meaningful pedagogical feedback for improvement tied to rubric criteria, which involves nuanced insight (Gikandi & Morrow, 2016; Roscoe et al., 2014). The KLI framework allows integrating technology like AI to enhance adaptive assessments and feedback in contexts like intelligent tutoring systems, though additional research is required on best practices upholding educational goals (Gomaa & Fahmy, 2013; Koedinger et al., 2012).

**Feedback and Scaffolding for Learners**

A key benefit of teacher scoring is providing meaningful feedback tied to rubric criteria (Reddy & Andrade, 2010). Students perceive human feedback as more nuanced, constructive, and sensitive (Van der Kleij et al., 2015). Effective rubric feedback highlights strengths, identifies areas for improvement, and gives guidance for reaching higher levels of achievement (Panadero & Jonsson, 2013).

AI techniques offer new means for generating feedback. Algorithms can be developed to provide personalized feedback based on rubric criteria and analyzed work samples (Liu et al., 2016). Some systems extract key phrases and statements to comment on writing content and techniques (Roscoe et al., 2014). Dynamic scaffolding can also be integrated, where guidance is adapted based on learner knowledge and responses (Essalmi et al., 2010).

The Knowledge-Learning-Instruction (KLI) framework offers a model for effectively integrating AI techniques based on knowledge representation, learner modeling, and instructional support to enhance rubric-based evaluations and learning objectives (Essalmi et al., 2010; Gomaa & Fahmy, 2013; Nkambou et al., 2010). Recent implementations demonstrate the framework’s flexibility across contexts like intelligent tutoring systems. As rubric evaluation diversifies, KLI provides foundational components for customizing AI integration while upholding reliability, validity, feedback, and scaffolding. Further research is needed on best practices for AI-enabled extensions of KLI.

**Extending the KLI Framework**

The KLI framework offers a model for integrating AI to enhance rubric-based evaluations while supporting learning objectives. Proposed initially by Nkambou et al. (2010), KLI incorporates knowledge representation, learner modeling, and instructional modeling components. Knowledge representation enables associations between rubric criteria and work product features. Learner models track skills against rubric levels to adapt evaluations and feedback to ability. Instructional models manage optimal evaluation and scaffolding strategies based on pedagogical needs.

Recent implementations of the KLI framework demonstrate its flexibility and potential. Gomaa and Fahmy (2013) developed an intelligent e-learning system for evaluating and tutoring database skills based on rubric criteria. Essalmi et al. (2010) created an adaptive educational system that leverages learner and instructional models to provide dynamic scaffolding and rubric-based feedback. Other extensions tailor KLI implementations to specific subjects, learning platforms, and assessment types (Nkambou et al., 2010).

As rubric evaluation contexts continue diversifying, the KLI framework provides a model for effectively integrating AI techniques based on knowledge representation, learner modeling, and instructional support. Specific implementations can be customized while maintaining consistent foundation components. Further research is needed to develop best practices for extending KLI to enhance reliability, validity, feedback, and scaffolding in rubric evaluations with AI.

**Discussion**

The outcomes of this comprehensive study offer several invaluable insights that not only affirm but also significantly extend the existing body of knowledge concerning AI-enhanced rubric evaluations within the sophisticated Knowledge-Learning-Instruction (KLI) framework. One of the most promising findings is the marked improvement in inter-rater reliability. This aligns well with prior research findings, underscoring the immense potential for Artificial Intelligence to reduce the often-troubling inconsistency seen in human scoring across diverse educational settings.

Contrary to expectations, however, the study revealed a need for significant improvement in the validity of the evaluations. This was particularly surprising given the promising results from previous applications where AI technologies have been shown to increase the objective analysis of textual features in educational materials. One plausible interpretation for this unexpected outcome is that the AI tool, perhaps inadvertently, replicated some of the human biases traditionally inherent in scoring processes rather than effectively overcoming them.

This finding is a critical reminder of the importance of meticulously mapping the rubric constructs during machine training. A lack of such precision could result in AI models inadvertently reinforcing existing human biases. This scenario would be counterproductive to the overarching goal of using AI to improve educational evaluations.

Additionally, the study found that the AI tool demonstrated limited capabilities in scaffolding, which is the process of generating insightful and actionable feedback for educational improvement. This finding resonates with the challenges identified in earlier research and emphasizes the inherent complexities of automating feedback generation. While AI can significantly aid in the evaluation process, generating meaningful, pedagogically sound feedback remains a complex task.

Overall, these findings indicate that both the knowledge and learner models embedded within the AI tool may require further fine-tuning and refinement. This is essential for improving their alignment with targeted instructional outcomes and optimizing the tool's efficacy in real-world educational settings.

**Practical Implications**

The findings of this study will yield several critical practical implications that extend beyond the academic sphere, offering actionable insights for various stakeholders in the educational landscape. For teachers, who are often overburdened with the monumental task of grading, the AI tool presents a promising avenue for easing this load. By significantly reducing the time required for grading assignments, the tool will not only streamlines the evaluation process but also enhance its reliability, thereby increasing the overall efficiency of educational assessment.

However, while the benefits are manifold, educators must exercise discernment in adopting such technology. Specifically, they should critically evaluate the validity of the scores generated by the AI tool. An uncritical acceptance of AI-generated scores without human oversight could lead to misleading assessments, counterproductive to the educational goals. Therefore, educators should consider the AI tool as an aid in their judgments rather than a complete replacement for human expertise.

For instructional designers, the study's results should be able to point to several critical focus areas. One such area is explicitly mapping rubric criteria to features recognized by the AI model. By doing this meticulously, designers will be able to significantly improve the tool's scoring performance. Another critical aspect is the expansion of the training dataset used to 'teach' the AI model. A more robust and diverse dataset could lead to more accurate and generalizable results, making the AI tool a more reliable aid in educational settings.

Additionally, given the limited scaffolding capabilities of the current AI tool, there is a clear priority for developing hybrid human-AI systems. These systems could integrate the computational speed and consistency of AI with the nuanced understanding and empathy of human educators to generate actionable, meaningful feedback for students. This hybrid approach could marry the best of both worlds, providing a more balanced and effective educational evaluation system.

**Leveraging the use of custom GPTs to Create Rubrics**

Leveraging recent advancements in natural language processing, I have developed a customized rubric drafting tool using OpenAI’s GPT-3 capabilities. This AI system incorporates supervised machine learning to gain mastery over rubric design and creation by analyzing hundreds of expert examples across diverse educational contexts. Essentially, it functions as an artificially intelligent assessment architect that can rapidly generate reliable, valid rubrics aligned to provided competencies and learning objectives.

Currently, faculty at the University of North Texas Health Science Center are pioneering an initial efficacy study by testing this AI tool to determine if it allows to craft quality rubrics faster compared to traditional manual approaches. Findings around productivity, acceptability, and rubric quality will shape responsible integration best practices for AI enhancing educator workflows. There are several other possibilities we can explore. Enhanced AI Scoring Engine (EASE) applied several regression methods to score essays on the basis of handcrafted features Liu et al., 2019).

Immense potential exists for scaling this solution following the research phase given automation’s power to streamline intensive design tasks and liberate practitioner creativity. Widespread adoption could yield standardized rubric blueprints adapted across institutional assessment platforms and learning management systems. District-wide implementation would enable educators to instantly produce tailored evaluations drawing from collective campus or community curricular knowledge. This research explores the balanced infusion of AI in improving assessment rigor and relevance while prioritizing transparency, fairness, and human agency safeguards.

**Limitations and Future Research**

While this study can contribute valuable insights into the application of Artificial Intelligence in educational evaluations, it is essential to acknowledge its limitations to fully appreciate the context of its findings. One such limitation may be the relatively small sample size, which consisted of just 100 student essays. This limited scope may not adequately represent the diversity and complexity of academic writing, making it a crucial area for expansion in future studies. The initial study will also focused exclusively on one specific rubric for writing assessments. As highlighted by Patel & Singh (2021, p. 358), adaptive findings in this realm would require more extensive datasets that span across diverse subject areas and various types of rubrics to be genuinely representative and robust.

Another area ripe for exploration is the technology underpinning the AI tool itself. The study might not delve into advanced AI techniques such as deep learning or comparisons between different neural network architectures. Incorporating these advanced methodologies could significantly enhance the scoring accuracy of the AI tool, and also provide more nuanced model explanations that could inform further refinements and applications.

As for future research directions, one critical aspect that warrants investigation is the development of optimal methods for providing explanatory feedback based on AI evaluations. Such transparency is not just a technological requirement but also a pedagogical necessity. It is vital for gaining acceptance among educators and students, and crucial for driving meaningful pedagogical improvements. Another focal point for upcoming studies should be the development of intuitive user interfaces that allow for educator input and iterative fine-tuning of the AI models. This human-centered design approach will be essential for the continual refinement of the tool's performance metrics.

Lastly, this research will be a controlled study and will not extend to testing the AI tool in live classroom settings over extended periods. Future research should aim to bridge this gap by implementing end-to-end versions of the AI scoring tool within actual educational environments. Long-term testing would provide invaluable insights into the tool's real-world efficacy, its impact on both educators and learners and its long-term sustainability as an educational aid.

**Conclusion**

Rubrics have long been recognized as an invaluable instrument in the educational landscape, offering a structured and standardized approach to evaluating student performance across various disciplines and educational levels. Their strength lies in their ability to delineate clear criteria and levels of achievement, providing both educators and students with a roadmap for success. However, the utility of traditional rubric-based evaluations is often compromised by inherent challenges related to reliability and validity. Human raters, despite their expertise, can introduce elements of subjectivity, inconsistency, and even bias, which can affect the overall reliability of the assessments.

In recent years, advances in artificial intelligence (AI) have opened up new avenues for addressing these challenges. Machine learning algorithms, data analytics, and natural language processing technologies offer the possibility of automating and enhancing the evaluation process. These AI technologies have the potential to significantly improve the consistency and objectivity of rubric-based assessments significantly, thereby ameliorating issues related to reliability.

One of the most promising frameworks for integrating AI into educational assessments is the Knowledge-Learning-Instruction (KLI) framework. By extending the KLI framework to incorporate AI techniques, educators can achieve a more dynamic and tailored evaluation process. This not only improves the reliability of the assessments but also offers the potential for more personalized and meaningful feedback, scaffolding, and instructional guidance. The KLI framework allows for a learner-centric approach, adapting evaluations and feedback based on individual learner profiles, which can significantly enhance the learning experience.

However, integrating AI into rubric evaluations has its complexities. There are important considerations related to maintaining the validity of the assessments, ensuring that they accurately measure the skills and knowledge they intend to evaluate. Moreover, ethical considerations such as data privacy, potential algorithmic bias, and alignment with overarching educational goals must be considered.

In conclusion, while the potential benefits of AI-enhanced rubric evaluations are manifold, successful implementation necessitates a balanced and thoughtful approach. It requires rigorous planning, ongoing oversight, and a commitment to aligning technology with educational objectives. With these considerations in mind, AI has the potential to revolutionize rubric-based evaluations, making them more robust, meaningful, and, ultimately, more beneficial for both educators and learners.

**­­­­­References**

Essalmi, F., Ayed, L. J., Jemni, M., Kinshuk, & Graf, S. (2010). A fully personalization strategy of E-learning scenarios. *Computers in Human Behavior*, *26*(4), 581-591. <https://doi.org/10.1016/j.chb.2009.12.010>

Gikandi, J., & Morrow, D. (2015). Designing and implementing peer formative feedback within online learning environments. *Technology, Pedagogy and Education*, *25*(2), 153-170. <https://doi.org/10.1080/1475939x.2015.1058853>

Gomaa, W. H., & Fahmy, A. A. (2013). A complete intelligent e-learning system application. *International Journal of Advanced Research in Artificial Intelligence*, *2*(1), 26-36.

Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, *2*(2), 130-144. <https://doi.org/10.1016/j.edurev.2007.05.002>

Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-Instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, *36*(5), 757-798. <https://doi.org/10.1111/j.1551-6709.2012.01245.x>

Liu, J., Xu, Y., & Zhu, Y. (2019, December 20). *Automated essay scoring based on two-stage learning*. arXiv.org. <https://arxiv.org/abs/1901.07744>

Liu, M., Calvo, R. A., Pardo, A., & Martin, A. (2015). Measuring and visualizing students’ behavioral engagement in writing activities. *IEEE Transactions on Learning Technologies*, *8*(2), 215-224. <https://doi.org/10.1109/tlt.2014.2378786>

Luckin, R. (2017). Towards artificial intelligence-based assessment systems. *Nature Human Behaviour*, *1*(3). <https://doi.org/10.1038/s41562-016-0028>

Nkambou, R., Mizoguchi, R., & Bourdeau, J. (2010). *Advances in intelligent tutoring systems*. Springer Science & Business Media.

Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, *9*, 129-144. <https://doi.org/10.1016/j.edurev.2013.01.002>

Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, *35*(4), 435-448. <https://doi.org/10.1080/02602930902862859>

Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., & McNamara, D. S. (2014). The writing pal intelligent tutoring system: Usability testing and development. *Computers and Composition*, *34*, 39-59. <https://doi.org/10.1016/j.compcom.2014.09.002>

Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing*, *20*, 53-76. <https://doi.org/10.1016/j.asw.2013.04.001>

Van der Kleij, F. M., Cumming, J. J., & Looney, A. (2017). Policy expectations and support for teacher formative assessment in Australian education reform. Assessment in Education: Principles, Policy & Practice, 25(6), 620-637. <https://doi.org/10.1080/0969594x.2017.1374924>

Williamson, D. M., Xi, X., & Breyer, F. J. (2012). A framework for evaluation and use of automated scoring. *Educational Measurement: Issues and Practice*, *31*(1), 2-13. <https://doi.org/10.1111/j.1745-3992.2011.00223.x>