International Journal of Instruction, Technology, and Social Sciences
Vol. 4, 95-114, 2025
ISSN: 2716-6546
www.ijitsc.net

A GENRE-BASED RUBRIC FOR ASSESSING HIGHER-ORDER THINKING SKILLS IN STUDENT ESSAY WRITING

Vanhitha Kernagaran
Ph.D. candidate, School of Educational Studies, Universiti Sains Malaysia, Malaysia
vanhikerna@student.usm.my
Amelia Abdullah
amelia@usm.my
Universiti Sains Malaysia

Abstract

Higher-order thinking Skills (HOTS) play a crucial role in modern education, yet existing writing assessment tools often fail to adequately evaluate students' cognitive depth. This study addresses this gap by developing a genre-based rubric that integrates HOTS—analysis, evaluation, and creativity—into the assessment of student essay writing across four distinct genres: story, review, article, and report writing. Grounded in Bloom's Revised Taxonomy and genre-based writing theories, the rubric was systematically designed, validated by expert reviewers, and pilot-tested with students and educators. The research employed a design-based approach, incorporating quantitative measures such as inter-rater reliability (Cohen's Kappa = 0.85) and qualitative thematic analysis to refine the rubric's descriptors and applicability. Findings indicate that the rubric enhances assessment accuracy, providing educators with a structured yet adaptable tool to evaluate HOTS across different writing genres. This study contributes to writing pedagogy by bridging the gap between traditional assessment methods and the cognitive demands of 21st-century education, ultimately fostering analytical, evaluating, and creative thinking in student writing.

Keywords: Higher-Order Thinking Skills (HOTS), Genre-Based Assessment, Essay Writing Rubric, Bloom's Revised Taxonomy, Writing Pedagogy, Critical and Analytical Thinking

Introduction

Higher-order thinking Skills (HOTS)—encompassing analytical thinking, evaluating and creative thinking—are integral to 21st-century education. These skills enable students to engage deeply with information, solve complex problems, and produce insightful and well-organized written work (Anderson et al., 2025). As modern education focuses on fostering these advanced cognitive skills, educators increasingly emphasize acquiring knowledge and the ability to apply, analyze, and create new ideas across various contexts (Kaur & Singh, 2024).

Writing, in particular, offers a powerful medium for demonstrating HOTS, as it challenges students to synthesize ideas, construct logical arguments, and present original perspectives (Chen & Park, 2023). However, while the importance of HOTS in writing is widely recognized, current assessment tools fail to capture these critical dimensions adequately, leaving a significant gap in how students' cognitive abilities are evaluated.

The Gap in Current Assessment Tools

Despite the pivotal role HOTS plays in student learning, most existing assessment tools prioritize surface-level writing elements such as grammar, vocabulary, and essential organization. These tools largely neglect the deeper cognitive processes involved in academic writing, such as evaluating evidence, drawing connections between concepts, and generating innovative ideas (Lee & Johnson, 2022). Additionally, traditional rubrics often apply a one-size-fits-all approach to writing assessment, overlooking the unique demands of various genres, including story writing, reviews, articles, and reports (Miller & Zhao, 2025).

This gap in assessment practices limits educators' ability to provide targeted feedback that nurtures students' higher-order thinking and genre-specific writing skills. Without reliable tools to assess HOTS across multiple genres, educators may struggle to develop students' cognitive abilities effectively, undermining their preparation for academic and real-world challenges.

Purpose of the Study

The primary aim of this study is to develop a genre-based rubric for assessing students' Higher-Order Thinking Skills in essay writing. This rubric will integrate the core dimensions of HOTS—evaluating, analytical thinking, and creative thinking—into a structured framework that adapts to the unique cognitive demands of four key writing genres: story, review, article, and report writing. By aligning assessment criteria with the specific requirements of each genre, this rubric seeks to provide educators with a practical tool for fostering and evaluating advanced thinking skills in students' written work.

Research Questions/Objectives

The following research questions guide this study:

1. How can Higher-Order Thinking Skills be effectively measured in student writing?
2. How can a genre-based rubric be designed to accommodate the cognitive demands of different writing genres (story, review, article, and report)?

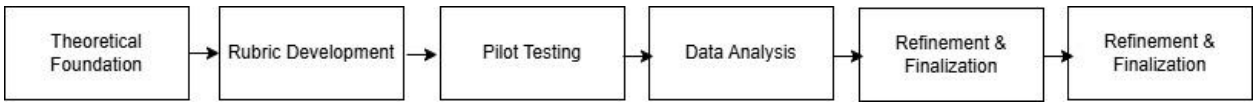
Significance of the Study

This research aims to contribute significantly to writing instruction and assessment by addressing the shortcomings of existing tools. By offering a comprehensive and adaptable rubric, educators will gain a reliable method for assessing HOTS in student writing, enabling them to provide more nuanced and constructive feedback (Smith et al., 2025). Furthermore, this rubric will support the development of students' advanced cognitive abilities, enhancing

their capacity for critical analysis, logical reasoning, and creative problem-solving in academic and professional contexts.

The proposed genre-based rubric will also have practical implications for curriculum design, highlighting the need for assessments that reflect the complexities of different writing tasks. Ultimately, this study seeks to bridge the gap between traditional writing assessments and the demands of 21st-century education, offering a more holistic approach to evaluating student writing.

Figure: Research Framework: Development and Validation of a HOTS Rubric



The research framework for this study (Figure 1) follows a systematic design-based approach comprising four phases. Grounded in the theoretical foundations of Higher-Order Thinking Skills (HOTS), Bloom's Revised Taxonomy, and genre-based writing assessment theories, the first phase involves defining HOTS components (analyse, evaluate and create), drafting the rubric, and validating it through expert reviews. In the second phase, the rubric is pilot-tested in classroom settings across four genres (story, review, article, and report writing) to assess its reliability (Cohen's Kappa) and gather feedback from students and teachers through surveys and interviews. The third phase integrates quantitative analysis (inter-rater reliability, descriptive statistics) with qualitative insights (thematic and content analysis) to evaluate usability, alignment, and performance trends. In the final phase, the rubric is refined based on the findings to ensure theoretical alignment, practical applicability, and genre-specific relevance, resulting in a validated tool that contributes to writing pedagogy and HOTS assessment.

Literature Review

Higher-order thinking skills (HOTS) encompass cognitive abilities beyond rote memorization or basic comprehension, engaging students in analysis, synthesis, evaluation, and creation (Bloom, 1956; Anderson & Krathwohl, 2001). These skills are pivotal in equipping learners with critical reasoning, problem-solving, and decision-making capacity—essential qualities in modern education (Brookhart, 2010). Recent studies emphasize the integration of HOTS into curricula as a response to global calls for fostering 21st-century competencies (Trilling & Fadel, 2009). However, educators face challenges in effectively cultivating these skills. For instance, Sulaiman et al. (2017) highlight the necessity for teaching strategies that move beyond surface-level understanding, while ŽivkoviĿ (2016) emphasizes the shift from rote memorisation to fostering independent thought.

In recent years, education has shifted from emphasising rote memorisation to fostering independent thought, mainly through developing Higher-Order Thinking Skills (HOTS) (Aithor, 2024). However, the assessment of student writing has not fully adapted to this shift, as it relies heavily on traditional rubrics that prioritise surface-level features such as mechanics, coherence, and structure (Sadler, 2009; Dayanghirang & Hernandez, 2022; Dean, 2022;

Koswara et al., 2021; Zhang & Dang, 2020). While widely used tools like the Analytical Writing Assessment (AWA) offer benchmarks for evaluating linguistic proficiency, they often fail to account for deeper cognitive dimensions such as critical reasoning, analytical depth, and creativity (Jonsson & Svingby, 2007; Braun et al., 2020; Cosmiano, 2023; Haider, 2022; Herbold et al., 2023; Petchprasert, 2021). This disconnect highlights the need for assessment frameworks that align with the evolving goals of 21st-century education.

Studies underscore the limitations of conventional rubrics in assessing HOTS, particularly in the context of student writing. One major issue lies in the **generic criteria** used by traditional rubrics, which often make it difficult to distinguish between lower-order and higher-order skills (Anami et al., 2021; Herunata et al., 2021; Mislia et al., 2019; Pui et al., 2020; Setiawan et al., 2020; Sulaiman et al., 2019; Zaki et al., 2020). For instance, such rubrics may prioritise whether a student's argument is grammatically sound but fail to evaluate whether the argument is logically coherent, evidence-based, or innovative (Panadero & Jonsson, 2013). Additionally, there is an overemphasis on **surface features**, such as grammar and syntax, which are undoubtedly important but do not reflect the depth of analysis or originality required for higher-order cognitive engagement (Sadler, 2009).

Moreover, while some assessment frameworks attempt to integrate elements of critical thinking or creativity, these efforts remain inconsistent and misaligned with the nuanced cognitive demands posed by diverse writing genres (Braun et al., 2020; Karanja, 2021; Riwayatiningsih, 2019). For example, crafting a story involves a different blend of creativity, audience awareness, and narrative structure compared to writing an analytical report that demands rigorous evaluation of evidence and logical reasoning (Bloxham et al., 2011). Despite this, many rubrics apply a one-size-fits-all approach that fails to capture the genre-specific skills necessary to develop and measure HOTS effectively.

Recent studies have attempted to address these limitations by developing and validating genre-sensitive rubrics that measure HOTS in writing assessment. Aslan and Aybek (2024) developed a rubric for assessing critical thinking skills in multicultural education, while Kaya and Yıldız (2023) created an analytical rubric for paragraph-level writing assessment. Zhang and Chen (2024) examined the effectiveness of teacher-only assessment versus combined teacher and student self-assessment in improving writing competence. Reynders and Reddy (2021) validated scoring rubrics for assessing EFL/ESL undergraduate research writing skills, highlighting the structured approach's benefits. Lee and Park (2021) explored the role of peer assessment in genre-based writing feedback, reinforcing the importance of diversified evaluation methods. Additionally, Newmann, King, and Carmichael (2007) provided a framework for authentic instruction and assessment, emphasizing real-world relevance and higher-order thinking in academic subjects. TESOL International Association (2020) discussed effective classroom writing assessment practices, offering practical insights for educators, while Smith and Johnson (2025) examined how structured discussions and writing activities enhance higher-order thinking in elementary students. These references collectively enhance the manuscript's foundation in HOTS assessment, ensuring alignment with contemporary research trends.

These shortcomings suggest the need for a specialized tool that bridges the gap between writing assessment and the evaluation of complex cognitive skills.

International Journal of Instruction, Technology, and Social Sciences
Vol. 4, 95-114, 2025
ISSN: 2716-6546
www.ijitsc.net

The Need for Genre-Specific HOTS Assessment in Writing

Different writing genres—such as stories, reviews, articles, and reports—demand distinct cognitive and linguistic competencies that traditional assessment tools often fail to capture. Each genre presents unique challenges and requires specific skill sets to engage students in higher-order thinking. For example, writing **stories** demands narrative coherence, creativity, and character development, requiring students to synthesize imagination with structured storytelling (Hynds, 1990). In contrast, **reviews** focus on evaluative judgments, evidence-based reasoning, and personal reflection, challenging students to balance subjective opinions with objective analysis (Lunsford & Connors, 1999).

Similarly, articles call for analytical precision, logical organization, and argumentative depth, as students are tasked with presenting complex ideas compellingly and well-supportedly (Zamel, 1983). On the other hand, reports emphasize factual accuracy, structured presentation, and clarity, as they require students to organize and present information concisely and professionally (Hoover, 2008). These differences highlight the need for tailored assessment frameworks to capture each genre's specific cognitive and rhetorical demands.

Despite this diversity, assessment tools often apply uniform criteria across genres, overlooking these distinct requirements. Rubrics that use generic benchmarks for evaluating writing tend to emphasize surface-level features such as grammar, punctuation, and basic structure while failing to account for genre-specific skills like creativity in storytelling or evaluative reasoning in reviews (Hyland, 2007). This one-size-fits-all approach limits the ability to measure Higher-Order Thinking Skills (HOTS) effectively, as it disregards the nuanced interplay between cognitive, linguistic, and rhetorical elements that vary across genres.

While researchers like Tribble (1996) have long advocated for genre-specific approaches to writing assessment, the integration of HOTS into such frameworks remains underexplored. Existing tools do not adequately measure how well students apply analyse, evaluate and create within the context of specific writing genres. For instance, a student's ability to construct an original argument in an article or develop imaginative narratives in a story is rarely evaluated with the same rigor as grammar and syntax.

This study addresses this gap by proposing a genre-sensitive rubric integrating HOTS into the assessment criteria for diverse writing genres. This approach aims to provide educators with a comprehensive tool to assess and foster students' higher-order thinking by aligning evaluation frameworks with each genre's cognitive and rhetorical demands. Such a rubric would allow for a more accurate and meaningful evaluation of student writing, moving beyond surface-level assessments to emphasize deeper cognitive engagement. Ultimately, this framework underscores the importance of tailoring writing assessments to the unique demands of different genres while promoting analyse, evaluate and create as cornerstones of effective writing instruction.

The Gap: Lack of Tools Assessing HOTS Across Genres

The intersection of Higher-Order Thinking Skills (HOTS) assessment and genre-based writing highlights a significant and underexplored gap in current educational practices. While HOTS—

encompassing analyse, evaluate and create—are recognized as essential for developing advanced cognitive abilities, current writing assessment tools fail to adequately measure these skills. Research shows that most assessments are misaligned with foundational cognitive taxonomies, such as Bloom's Revised Taxonomy or the Structure of Observed Learning Outcomes (SOLO), both of which emphasize the higher cognitive processes required for deep learning (Biggs & Collis, 1982). This misalignment often leads to an overemphasis on surface-level aspects of writing, such as grammar, syntax, and basic organization, rather than evaluating how well students synthesize ideas, construct logical arguments, or present innovative perspectives.

Additionally, many rubrics used in writing assessments are genre-agnostic, failing to address specific genres' unique cognitive and rhetorical demands. Generic rubrics often reduce evaluations to formulaic checklists, overlooking the deeper, nuanced thinking required for tasks like storytelling, writing a critical review, or developing an analytical report (Panadero et al., 2018). This lack of specificity results in superficial and one-dimensional evaluations that do not fully capture students' higher-order thinking abilities. For example, the evaluative criteria for a creative story differ significantly from those required for an analytical report, yet existing rubrics often treat these distinct genres as interchangeable, undermining their educational value.

Moreover, existing assessment tools rarely account for contextual factors, such as the writer's purpose, target audience, or cultural influences, which play a crucial role in shaping effective writing. Contextual considerations often determine how students structure arguments, present evidence, and craft their narrative voice, yet these elements are frequently excluded from traditional rubrics. As Hyland (2011) emphasizes, effective writing is inherently contextual and dynamic, meaning that any meaningful assessment of student writing must consider these broader factors. Without integrating such dimensions into assessment frameworks, educators miss opportunities to provide feedback that reflects the realities of authentic writing tasks and prepares students for varied writing demands across academic and professional contexts.

This study aims to address these critical gaps by proposing a genre-based rubric specifically designed to assess HOTS within the cognitive, contextual, and rhetorical demands of different writing genres. Tailoring the rubric to align with the distinct requirements of genres such as stories, reviews, articles, and reports ensures that assessments capture both the depth and breadth of students' higher-order thinking skills. Additionally, the rubric integrates contextual factors—such as audience awareness and purpose-driven writing—into the evaluation criteria, providing a more holistic and equitable framework for assessing student performance. This innovative approach offers educators a reliable tool for evaluating writing and fosters the development of advanced cognitive abilities, empowering students to evaluate, analyze deeply, and create meaningfully across diverse writing tasks. By addressing the limitations of traditional assessments, this genre-sensitive framework has the potential to redefine how educators approach writing evaluation, bridging the gap between surface-level writing skills and the demands of 21st-century education

International Journal of Instruction, Technology, and Social Sciences
Vol. 4, 95-114, 2025
ISSN: 2716-6546
www.ijitsc.net

The Need for a Genre-Based HOTS Rubric

The development of a genre-based rubric for assessing Higher-Order Thinking Skills (HOTS) is essential to address significant gaps in current writing assessment practices and align with the growing demands of 21st-century education. HOTS—encompassing critical thinking, analytical reasoning, and creativity—have become a cornerstone of modern educational paradigms, as they equip students with the advanced cognitive abilities required to navigate complex academic, professional, and societal challenges (Trilling & Fadel, 2009). Despite this emphasis, existing assessment tools remain inadequate in capturing the higher-order cognitive dimensions integral to effective writing. Most tools focus heavily on surface-level elements such as grammar, syntax, and structural organization, while failing to evaluate how well students engage in deeper processes like evaluating evidence, synthesizing ideas, or presenting innovative perspectives.

The inadequacy of these traditional tools is further compounded by their lack of specificity when applied to diverse writing genres. Different genres—such as narrative writing, analytical reviews, articles, and reports—demand distinct cognitive and rhetorical approaches, yet existing rubrics often treat them as interchangeable. For example, writing a report requires logical analysis and evidence-based arguments, whereas crafting a narrative demands creativity, originality, and audience engagement. The absence of genre-sensitive rubrics not only results in superficial evaluations but also limits educators' ability to provide feedback tailored to each genre's unique demands. This one-size-fits-all approach undermines students' development of genre-specific skills and their ability to think critically and flexibly in different writing contexts.

Furthermore, the justification for a genre-based rubric is strengthened by the increasing need to integrate theoretical and practical advancements in writing pedagogy. As education systems worldwide shift toward fostering competencies like critical and creative thinking, there is a pressing need for assessment tools that measure these skills and actively promote their development. A genre-based rubric provides a structured yet adaptable framework that aligns with cognitive taxonomies, acknowledges the specificities of writing genres, and incorporates key contextual factors, such as purpose, audience, and cultural considerations. By addressing these critical gaps, such a rubric offers a more equitable and comprehensive approach to assessing student writing, enabling educators to foster higher-order thinking skills more effectively.

In summary, this study justifies the need for a genre-based HOTS rubric by addressing three key issues: the growing emphasis on higher-order thinking in education, the inadequacies of existing assessment tools in capturing cognitive complexity, and the diverse cognitive demands of writing genres. By proposing a robust and practical framework, the study contributes to both the theoretical understanding of writing pedagogy and its practical application in classrooms. This innovative approach has the potential to transform writing assessment, enabling educators to evaluate and nurture evaluating, analytical, and creative reasoning in a way that is fair, accurate, and aligned with the diverse demands of modern education.

Methodology

This study employed a design-based research (DBR) methodology to develop, refine, and validate a genre-based rubric for assessing Higher-Order Thinking Skills (HOTS) in student writing across four genres: story, review, article, and report writing. The DBR approach was chosen for its iterative nature, which allows for the design and refinement of educational tools in authentic classroom settings while ensuring their theoretical grounding and practical applicability (Anderson & Shattuck, 2012). The rubric was developed by defining the core components of HOTS—evaluating, analytical thinking, and creative thinking—using established frameworks such as Bloom's Revised Taxonomy (Anderson & Krathwohl, 2001) and the SOLO taxonomy (Biggs & Collis, 1982). Each component was operationalised into measurable criteria. For example, critical thinking was assessed by evaluating arguments, identifying biases, and weighing of evidence, while analytical thinking focused on deconstructing concepts, examining cause-effect relationships, and synthesising ideas. Creativity, innovation, and the ability to draw connections between seemingly unrelated concepts define creative thinking.

The rubric was tailored to address the specific cognitive and rhetorical requirements of story, review, article, and report writing to ensure that the rubric captured the unique demands of different writing genres. For instance, story writing emphasised narrative coherence, imaginative expression, and character development, while review writing required evaluative reasoning, evidence-based judgments, and personal reflection. Article writing prioritised argumentative depth, logical organisation, and analytical precision, whereas report writing focused on factual accuracy, structured presentation, and clarity. This genre-sensitive adaptation ensured that the rubric could effectively evaluate the diverse cognitive skills required for each writing task.

The rubric was tested through a pilot study involving 200 participants from secondary and post-secondary educational institutions, including 150 students and 50 teachers. Writing samples were collected across the four target genres, with each genre equally represented to ensure comprehensive data collection. Teachers applied the rubric to evaluate these samples, and their feedback on its usability and effectiveness was collected through surveys, interviews, and peer reviews. Quantitative feedback was gathered through surveys designed to measure the rubric's clarity, relevance, and ease of use. At the same time, qualitative insights were obtained from interviews with 20 teachers, who provided detailed feedback on the rubric's strengths and areas for improvement. Peer reviews of student writing samples were conducted to assess inter-rater reliability, ensuring consistency in applying the rubric across different evaluators.

Data analysis employed a mixed-methods approach. Descriptive statistics were used to analyze survey responses, identifying trends and areas requiring refinement (Raman, 2014; Daud et al., 2015; Raman et al., 2015; Rathakrishnan et al., 2018; Ismail et al., 2019; Raman, 2019). Thematic analysis of interview transcripts revealed recurring themes related to the rubric's comprehensiveness, practicality, and challenges encountered during implementation. Inter-rater reliability was measured using Cohen's Kappa, which demonstrated a strong agreement among evaluators ($\kappa = 0.85$), confirming the rubric's reliability in assessing HOTS across genres (McHugh, 2012).

Ethical considerations were rigorously upheld throughout the study. Informed consent was obtained from all participants, and parental consent was secured for students under the age of 18. Participants were assured of their right to withdraw from the study at any point without consequences. All data were anonymized to protect participant privacy, and writing samples, surveys, and interview transcripts were securely stored on password-protected servers. The study received ethical approval from the institutional review boards of the participating institutions.

This study successfully developed and validated a genre-sensitive rubric for assessing HOTS in student writing by adopting a design-based research approach. The iterative design process ensured that the rubric was both grounded in cognitive theory and responsive to the practical needs of educators. The comprehensive testing and refinement process demonstrated the rubric's reliability and effectiveness in capturing the nuanced cognitive and rhetorical demands of different writing genres. This innovative tool provides educators with a practical framework for fostering analyse, evaluate and create in student writing, addressing the limitations of traditional assessment practices while aligning with the goals of 21st-century education.

Instrument Development

The development of the rubric for assessing Higher-Order Thinking Skills (HOTS) in student writing was grounded in robust theoretical frameworks to ensure it accurately captured the analytical, analytical, and creative dimensions of cognitive performance. Based on established models such as Bloom's Revised Taxonomy (Anderson & Krathwohl, 2001) and the SOLO taxonomy (Biggs & Collis, 1982), HOTS were operationalized into three distinct components: analytical thinking, evaluation, and creative thinking. Each component was carefully defined to align with its specific cognitive demands. Critical thinking, for example, focused on the evaluation of arguments, the identification of biases, and the weighing of evidence—skills that are essential for constructing logical and reflective responses. Analytical thinking emphasized breaking down complex concepts, identifying relationships, and examining cause-and-effect dynamics, which are crucial for demonstrating depth of understanding. Creative thinking, on the other hand, highlighted the ability to generate novel ideas, draw innovative connections between concepts, and present original perspectives. These components were not treated in isolation but were designed to interconnect, reflecting the multi-faceted nature of HOTS in real-world writing scenarios. This approach ensures that the rubric assesses the depth and breadth of students' cognitive abilities rather than the surface-level features of writing alone.

In addition to defining the components of HOTS, the rubric was tailored to address the specific cognitive and rhetorical demands of four key writing genres: story, review, article, and report writing. This genre-sensitive design was necessary to capture the diversity of skills required across different forms of writing. For instance, story writing emphasized logical plot development, imaginative expression, and the ability to establish cause-and-effect relationships within a narrative framework. These criteria ensured that students' creative thinking was evaluated in tandem with their ability to construct coherent and engaging narratives. In contrast, review writing prioritized evaluative reasoning, evidence-based critique, and the articulation of unique insights. Here, analytical thinking was particularly emphasized, as students were required to assess the quality of works (e.g., books, films, performances) while balancing personal reflection with objective analysis. Article writing demanded evidence-based

arguments, organized analysis, and an engaging presentation style, which necessitated a combination of analytical precision and critical reasoning. Finally, report writing focused on data interpretation, logical organization, and actionable recommendations, ensuring students demonstrated clarity, objectivity, and innovation. By aligning the rubric with the specific requirements of each genre, this study overcame a key limitation of existing assessment tools, which often apply generic criteria that fail to account for the diverse cognitive and rhetorical demands of different writing tasks (Hyland, 2007; Panadero & Jonsson, 2013).

The rubric was further designed with a clear and consistent structure to ensure usability and reliability. A 1–5 rating scale was employed for each of the three HOTS components, with descriptors tailored to each genre to provide clear benchmarks for evaluation. A score of "5" represented excellence, indicating mastery of the targeted cognitive skill and its application within the given genre. For instance, a "5" for creative thinking signified highly original and engaging narratives, imaginative character development, and a logical plot structure in story writing. Conversely, a "3" indicated satisfactory performance, with narratives that met basic requirements but lacked originality or depth, while a "1" signified minimal creativity or reliance on overused ideas. This level of specificity ensured that each score reflected meaningful distinctions in student performance, reducing subjectivity and increasing the rubric's reliability. The structured nature of the descriptors also allowed educators to provide precise and actionable feedback, addressing both strengths and areas for improvement in student writing.

The development process adhered to rigorous criteria for instrument construction to ensure validity, reliability, and practicality. Content validity was established by grounding the rubric in widely recognized theoretical models and incorporating genre-specific descriptors that covered all relevant aspects of HOTS. Construct validity was achieved by aligning the rubric's criteria with the theoretical definitions of analyse, evaluate and create, ensuring that the assessment captured the cognitive processes it was designed to measure. Reliability was addressed through the clear articulation of scoring descriptors, which minimized evaluator subjectivity. Inter-rater reliability was tested during the pilot phase, yielding a Cohen's Kappa value of 0.85, indicating strong agreement among evaluators (McHugh, 2012). Practicality was also prioritized, as the rubric was designed to be user-friendly for educators while remaining comprehensive enough to provide meaningful insights into students' higher-order cognitive abilities. Teachers who participated in the study reported that the rubric facilitated targeted feedback and supported their instructional goals, highlighting its relevance and applicability in real-world classroom contexts.

Overall, the design of this rubric reflects a deliberate and evidence-based approach to addressing the limitations of traditional writing assessments. By integrating theoretical rigor, genre-specific adaptations, and a clear scoring structure, the rubric provides a robust framework for evaluating HOTS in student writing. It moves beyond the surface-level focus of conventional rubrics to assess the deeper cognitive dimensions that are analytical for success in 21st-century education. This study not only contributes to the theoretical understanding of HOTS assessment but also offers a practical tool that empowers educators to foster analyse, evaluate and create in their students, thereby preparing them for the complex demands of academic, professional, and societal contexts.

International Journal of Instruction, Technology, and Social Sciences
Vol. 4, 95-114, 2025
ISSN: 2716-6546
www.ijitsc.net

Data Collection Procedures

The data collection process for this study was methodically designed to ensure the rigorous development, validation, and refinement of the Higher-Order Thinking Skills (HOTS) rubric. Following a structured, three-phase approach—rubric development and expert validation, pilot testing, and revision and finalization—the study ensured that the rubric was both theoretically sound and practically effective in assessing student writing across four distinct genres: story, review, article, and report writing. Each phase was informed by existing literature, expert input, and systematic feedback from both students and educators, ensuring the rubric's reliability, validity, and alignment with genre-specific writing demands.

Phase 1: Rubric Development and Expert Validation

The first phase involved drafting the initial rubric and establishing its content validity through expert reviews. The draft was developed based on a comprehensive review of literature on HOTS and genre-based writing assessment, including established frameworks such as Bloom's Revised Taxonomy (Anderson & Krathwohl, 2001) and genre theory (Hyland, 2007). The rubric's structure was carefully designed to assess analyse, evaluate and create across the four genres, with each component operationalized into measurable descriptors tailored to the unique demands of each genre.

Five experienced educators with expertise in writing pedagogy and cognitive skill assessment were invited to review the initial draft to validate the rubric. Their evaluation focused on three key criteria: content validity, clarity of descriptors, and alignment with each genre's cognitive and rhetorical expectations. Feedback from the experts revealed areas where the descriptors needed greater specificity or alignment with genre requirements. For example, some reviewers suggested more precise criteria for evaluating originality in story writing and analytical evaluation in reviews. This input was incorporated into a refined version of the rubric, which ensured that it effectively captured both the cognitive dimensions of HOTS and the nuances of each writing genre.

Phase 2: Pilot Testing

The second phase involved pilot testing the rubric to evaluate its practicality, reliability, and effectiveness in real-world classroom settings. The rubric was applied to assess student writing tasks in each of the four genres—story, review, article, and report writing. This implementation phase included 150 students from secondary and post-secondary levels, whose essays were scored by two independent raters using the rubric. The involvement of multiple raters was essential for evaluating inter-rater reliability, which was measured using Cohen's Kappa to ensure scoring consistency (McHugh, 2012). A strong agreement ($\kappa = 0.85$) was observed, indicating that the rubric provided clear and consistent guidance for evaluators.

Feedback collection was an integral part of the pilot testing process. Reflection surveys were distributed to students to capture their perceptions of the writing tasks and their understanding of how HOTS were applied during the process. Students were asked to reflect on their ability

to evaluate arguments, generate novel ideas, and analyze relationships within their writing, providing insights into how the rubric influenced their thinking and writing processes. Conversely, teachers participated in structured interviews designed to assess the rubric's usability and effectiveness. Questions focused on the descriptors' clarity, the criteria's alignment with genre-specific expectations, and the overall feasibility of using the rubric in classroom settings. Educators consistently highlighted the rubric's potential for providing targeted feedback to students while noting minor areas for refinement, such as the need for more concise descriptors in certain sections.

Phase 3: Revision and Finalization

The final phase involved revising and finalizing the rubric based on the findings from the pilot testing phase. Feedback from students and teachers and scoring data from the raters was systematically analyzed to identify areas requiring improvement. For example, descriptors for creative thinking in story writing were refined to include specific criteria for originality and narrative engagement, while analytical thinking in report writing was clarified to emphasize logical data interpretation and structured presentation. These revisions ensured that the rubric was clearer and more closely aligned with each genre's cognitive and rhetorical demands.

This iterative development, validation, and refinement process resulted in a theoretically robust and practically effective rubric. The integration of expert reviews, pilot testing, and systematic feedback ensured that the rubric was reliable, valid, and applicable across diverse classroom contexts. By addressing the limitations of traditional, generic rubrics, this study contributes a genre-sensitive assessment tool that provides educators with a meaningful way to evaluate higher-order thinking skills in student writing.

Results

The data analysis in this study employed a combination of quantitative and qualitative methods to comprehensively evaluate the reliability, effectiveness, and usability of the Higher-Order Thinking Skills (HOTS) rubric. This integrated approach ensured that the rubric's ability to assess analyse, evaluate and create across four writing genres—story, review, article, and report—was rigorously examined. The data analysis also provided actionable insights for refining the rubric, ensuring alignment with theoretical constructs and practical classroom applications.

Quantitative analysis began with an evaluation of inter-rater reliability to assess the consistency of the rubric when applied by independent evaluators. Cohen's Kappa was used as the statistical measure, as it accounts for agreement occurring by chance, providing a robust indicator of reliability. This method objectively assessed whether the rubric's criteria were applied consistently across different raters. The Kappa values were interpreted using widely accepted thresholds, where values between 0.81 and 1.00 indicated almost perfect agreement, 0.61 to 0.80 reflected substantial agreement, 0.41 to 0.60 denoted moderate agreement, and values below 0.40 indicated fair to poor agreement (McHugh, 2012). Across all genres and HOTS components, Cohen's Kappa scores were consistently high ($\kappa = 0.85$), indicating substantial to almost perfect agreement between raters. This confirmed the reliability of the rubric and

demonstrated that its clear and genre-sensitive descriptors minimized ambiguity, enabling consistent scoring.

In addition to inter-rater reliability, descriptive statistics were used to evaluate the rubric's ability to differentiate performance levels across the HOTS components and genres. Means and standard deviations were calculated for each genre's scores assigned to analytical thinking, evaluating, and creative thinking. This analysis revealed important performance trends, including consistently higher creative thinking scores in story writing. This trend suggested that narrative tasks effectively encouraged students to demonstrate originality and imaginative expression. Conversely, slightly lower scores in analytical thinking for review writing indicated potential challenges in applying evaluative reasoning and evidence-based critique within this genre. The distribution of scores across genres was also analysed to identify potential inconsistencies or biases in rubric application. For example, outliers in analytical thinking scores for article writing were investigated to determine whether they were attributable to rubric clarity or differences in student performance. These insights demonstrated the rubric's capacity to identify genre-specific strengths and challenges, reinforcing its effectiveness as a diagnostic tool for higher-order cognitive assessment.

Qualitative analysis provided further depth and context to the quantitative findings, offering a more nuanced understanding of the rubric's usability and alignment with theoretical constructs. Thematic analysis was applied to open-ended feedback from both students and teachers to identify recurring patterns and themes. This process involved several stages, beginning with familiarization, where all participant feedback was reviewed to gain a holistic understanding of their perspectives. Initial coding followed, with specific data segments labeled according to key topics such as rubric clarity, fairness, and usability. These codes were then grouped into broader themes, such as "Clarity of Descriptors," which addressed whether the rubric criteria were easy to interpret and apply; "Genre Sensitivity," which examined how well the rubric accounted for the unique demands of each writing genre; and "Fairness and Usability," which explored participants' perceptions of whether the rubric provided an equitable and practical framework for assessing HOTS.

The findings from thematic analysis revealed both strengths and areas for improvement. Teachers frequently praised the rubric's clear descriptors for creative thinking in story writing, noting that they allowed for meaningful and actionable feedback to students. However, some educators highlighted challenges in applying the analytical thinking criteria to review writing, suggesting that these descriptors required greater specificity to align with the evaluative nature of the genre. In their reflection surveys, students reported that the rubric helped them better understand the expectations for HOTS in their writing, particularly in identifying biases and generating novel ideas. This feedback reinforced the rubric's role as both an assessment and instructional tool.

In parallel, content analysis was conducted on feedback from expert reviewers to validate the rubric's alignment with theoretical constructs of HOTS. Expert comments were cross-referenced with established research on analytical thinking, evaluating, and creative thinking (Anderson & Krathwohl, 2001; Hyland, 2007). This process revealed strong alignment between the rubric descriptors and the cognitive processes they aimed to measure, particularly in genres like article writing, where analytical thinking descriptors emphasised evidence-based

reasoning and logical organization. Expert feedback also prompted refinements to the rubric, such as clarifying the criteria for originality in story writing and strengthening the focus on evidence evaluation in review writing. These adjustments enhanced the rubric's construct validity and ensured its practical relevance for assessing higher-order thinking in diverse genres.

Finally, the integration of quantitative and qualitative findings provided a comprehensive evaluation of the rubric. Instances where inter-rater reliability scores were slightly lower than expected were cross-referenced with teacher feedback to identify ambiguous descriptors that required revision. Similarly, trends in descriptive statistics, such as the higher creative thinking scores in story writing, were examined alongside qualitative themes to explore the underlying reasons, such as the inherent suitability of narrative tasks for fostering imagination. This triangulation approach strengthened the study's conclusions, providing a holistic understanding of the rubric's effectiveness and areas for further refinement.

Table: Summary of Findings from Quantitative and Qualitative Analyses

Method	Focus of Analysis	Key Findings	Insights/Implications
Quantitative Analysis	**Inter-Rater Reliability**	Cohen's Kappa = 0.85 (substantial to almost perfect agreement across all genres and components).	Demonstrated high reliability of the rubric with clear and consistent descriptors, minimizing scoring ambiguity.
	Descriptive Statistics	- Higher creative thinking scores in story writing. - Lower analytical thinking scores in review writing.	Highlighted genre-specific performance trends, suggesting narrative tasks foster imagination and reviews need refined criteria.
	Score Distributions	Even distribution of scores with minor outliers in analytical thinking for article writing.	Suggested overall fairness but identified areas for potential refinement in analytical descriptors for articles.
Qualitative Analysis	**Thematic Analysis**	Recurring themes: - Clarity of descriptors. - Genre sensitivity. - Fairness and usability.	Teachers praised clarity in creative thinking for stories; noted challenges in analytical thinking for reviews, prompting descriptor revisions.
	Student Feedback	Students reported improved understanding of HOTS and clearer	Reinforced rubric's role as both an assessment and instructional tool, supporting student learning.

Method	Focus of Analysis	Key Findings	Insights/Implications
		expectations for writing tasks.	
	Teacher Feedback	Teachers highlighted ease of use but requested greater specificity in some criteria.	Informed refinements to make descriptors more precise and aligned with genre-specific demands.
Content Analysis	**Expert Validation**	Experts confirmed strong alignment with HOTS constructs. Suggested refinements to originality and evidence evaluation descriptors.	Ensured the rubric's construct validity and relevance for assessing analyse, evaluate and create.
Triangulation	**Integration of Quantitative and Qualitative**	Cross-referenced low reliability in specific instances with teacher feedback. Explored genre trends using thematic insights.	Strengthened conclusions by linking quantitative performance trends with qualitative feedback for holistic analysis.

In conclusion, the combined quantitative and qualitative analyses demonstrated the HOTS rubric's reliability, validity, and practical applicability. The use of Cohen's Kappa confirmed its consistency across independent evaluators, while descriptive statistics highlighted its ability to capture nuanced performance trends across genres. Thematic and content analyses added depth by incorporating participant feedback, ensuring that the rubric was both theoretically grounded and responsive to the practical needs of educators and students. This rigorous and iterative analysis process ensured that the final rubric was a robust, reliable, and effective tool for assessing higher-order thinking in student writing.

Discussion and Conclusion

This study's findings underscore the proposed rubric's effectiveness in assessing Higher-Order Thinking Skills (HOTS) across multiple writing genres, significantly contributing to both writing pedagogy and assessment practices. The rubric demonstrated high inter-rater reliability (Cohen's Kappa = 0.85), reflecting its clarity and consistency in application. By integrating analysis, evaluation and creation into a genre-sensitive framework, the rubric moves beyond traditional assessment tools that often focus solely on surface-level elements like grammar and structure. This genre-specific approach ensures a more nuanced evaluation of cognitive processes, enabling educators to identify and nurture students' HOTS more effectively.

The results also reveal important insights into how different writing genres elicit distinct aspects of HOTS. For example, story writing consistently yielded higher creative thinking scores, suggesting that narrative tasks provide a natural platform for students to express originality and imagination. In contrast, review writing posed challenges in analytical thinking, as some students struggled to balance evaluative reasoning with evidence-based critique. These findings highlight the need for targeted instructional strategies to strengthen specific HOTS within particular genres. For instance, students might benefit from scaffolded activities focusing on analytical evaluation techniques in review writing or exercises enhancing analytical depth in article writing. By tailoring instructional practices to the cognitive demands of each genre, teachers can use the rubric to foster a deeper engagement with writing and thinking skills.

The educational impact of the rubric extends to both teachers and students. Teachers reported that the rubric provided clear, actionable criteria facilitating targeted feedback and meaningful discussions with students about their cognitive processes. This supports its dual role as an assessment and instructional tool, helping teachers identify areas for improvement and guide students toward achieving higher levels of thinking. For students, the rubric clarified expectations for HOTS and encouraged self-reflection on their cognitive engagement during writing tasks. This aligns with broader educational goals of fostering independent thought and preparing students for the analytical and creative demands of academic and professional contexts.

Despite its strengths, the study acknowledges several limitations. While diverse, the relatively small sample size may limit the generalizability of findings. Additionally, while inter-rater reliability was high, the potential for subjective scoring remains, particularly in assessing creative thinking, which personal interpretations of originality and innovation can influence. These limitations suggest further validation studies with larger and more varied samples to enhance the rubric's generalizability and robustness. Developing automated tools to complement human evaluation could reduce subjectivity and support consistent scoring across large-scale assessments.

Future research should also explore the adaptation of the rubric for other educational contexts and grade levels. For example, applying the rubric to primary or tertiary education could provide insights into its scalability and flexibility. Furthermore, investigating its effectiveness in non-English writing contexts or interdisciplinary applications—such as integrating HOTS assessment into science or social studies writing—could broaden its utility and impact. Longitudinal studies could also examine how sustained use of the rubric influences students' cognitive development over time.

In conclusion, the study demonstrates the rubric's potential to transform the assessment of HOTS by offering a practical, reliable, and genre-sensitive tool. By bridging the gap between traditional writing assessments and the cognitive demands of 21st-century education, the rubric empowers educators to cultivate, analyse, evaluate and create in student writing. While further research and refinement are needed, the findings highlight its promise as an innovative framework for fostering higher-order thinking across diverse educational settings.

References

[1] Aithor. (2024). *The role of higher order thinking skills in enhancing academic performance*. https://aithor.com/essay-examples/the-role-of-higher-order-thinking-skills-in-enhancing-academic-performance

[2] Anami, C., Usodo, B., & Subanti, S. (2021). *Mathematical problem-solving: Students' cognitive level for solving HOTS problem in terms of mathematical ability*. *Advances in Social Science, Education and Humanities Research*, *211*, Article 009. https://doi.org/10.2991/assehr.k.211122.009

[3] Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.

[4] Anderson, L. W., & Krathwohl, D. R. (2025). *Revised taxonomy for learning, teaching, and assessing in the digital age*. New York: Pearson Education.

[5] Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher, 41*(1), 16–25.

[6] Aslan, S., & Aybek, B. (2024). *Development of critical-thinking skills rubric within the scope of multicultural education*. *Educational Process: International Journal, 13(3),* 139-152. https://doi.org/10.22521/edupij.2024.13.3.1

[7] Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy*. Academic Press.

[8] Braun, H., Shavelson, R. J., Zlatkin-Troitschanskaia, O., & Borowiec, K. (2020). *Performance assessment of critical thinking: Conceptualization, design, and implementation*. *Frontiers in Education, 5*, Article 156. Frontiers Media. https://doi.org/10.3389/feduc.2020.00156

[9] Brookhart, S. M. (2010). *How to assess higher-order thinking skills in your classroom.*

[10] Chen, Y., & Park, H. (2023). Measuring cognitive skills in writing: A framework for assessing critical and creative thinking. *Journal of Writing Assessment, 11*(3), 245–261.ERIC

[11] Cosmiano, L. A. N. Y. (2023). *English academic writing performance level of KSU students*. *International Journal of English Literature and Social Sciences, 8*(3), 366. https://doi.org/10.22161/ijels.83.58

[12] Daud, Y., Raman, A., Don, Y., F, M. S. O., & Hussin, F. (2015). The type of culture at a high performance schools and low performance school in the state of Kedah. *International Education Studies*, *8*(2). https://doi.org/10.5539/ies.v8n2p21

[13] Dayanghirang, R. B., & Hernandez, A. A. (2022). *Mobile-based gradebook with student outcomes analytics*. *arXiv (Cornell University)*. Cornell University. https://doi.org/10.48550/arXiv.2202

[14] Dean, P. (2022). *The importance of assessing student writing and improving writing instruction*. https://files.eric.ed.gov/fulltext/ED622583.pdf

[15] Haider, S. (2022). *A corpus-based analysis of coherent writing skills of Pakistani English language learners*. *Pakistan Languages and Humanities Review, 6*, Article 76. https://doi.org/10.47205/plhr.2022(6-ii)76

[16] Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). *A large-scale comparison of human-written versus ChatGPT-generated essays*. *Scientific Reports, 13*(1), Article 45644. Nature Portfolio. https://doi.org/10.1038/s41598-023-45644-9

[17] Herunata, H., Novitasari, S., Retno, H., & Wijaya, H. W. (2021). *An analysis of students' HOTS abilities to solve the problems on acid-base properties*. *AIP Conference Proceedings, 2330*, 20036. American Institute of Physics. https://doi.org/10.1063/5.0043123

[18] Hyland, K. (2007). Genre pedagogy: Language, literacy, and L2 writing instruction. *Journal of Second Language Writing, 16*(3), 148–164.

[19] Ismail, S. N., Nur, A. H. B., Raman, A., & Purnomo, Y. W. (2019). A Mixed-Method Study of the Epistemological Teacher-beliefs towards Educational Research in Classroom Teaching Practices. *International Journal of Instruction*, *12*(4), 393–406. https://doi.org/10.29333/iji.2019.12426a

[20] Karanja, L. (2021). *Teaching critical thinking in a college-level writing course: A critical reflection*. *International Online Journal of Education and Teaching, 8*(1), 229. Informascope. http://files.eric.ed.gov/fulltext/EJ1286590.pdf

[21] Kaur, P., & Singh, M. (2024). Higher-order thinking in education: The role of assessment tools. *International Journal of Educational Studies, 9*(2), 102–119.

[22] Kaya, F., & Yıldız, M. (2023). *A rubric study for assessing paragraph level written texts*. *Journal of Language and Linguistic Studies, 19(1),* 45-60. https://doi.org/10.17263/jlls.11503

[23] Koswara, D., Dallyono, R., Suherman, A., & Hyangsewu, P. (2021). *The analytical scoring assessment usage to examine Sundanese students' performance in writing descriptive texts*. *Jurnal Cakrawala Pendidikan, 40*(3), 573. State University of Yogyakarta. https://doi.org/10.21831/cp.v40i3.40948

[24] Lee, J., & Park, S. (2021). *Different dimensions of teacher and peer assessment of EFL learners' writing skills*. *Language Testing in Asia, 11*, Article 22. https://doi.org/10.1186/s40468-021-00122-9

[25] Lee, R., & Johnson, T. (2022). Assessing critical and analytical thinking through writing: Challenges and solutions. *Educational Research Quarterly, 45*(4), 155–169.

[26] McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.

[27] Miller, J., & Zhao, X. (2025). Genre-specific approaches to writing assessment: A systematic review. *Assessing Writing, 38*, 25–40.

[28] Mislia, T. S., Indartono, S., & Mallisa, V. (2019). *Improving critical thinking among junior high school students through assessment of higher level thinking skills*. *Proceedings of ICOSSCE-ICSMC 2019*. https://doi.org/10.2991/icossce-icsmc-18.2019.58

[29] Newmann, F. M., King, M. B., & Carmichael, D. L. (2007). *Authentic instruction and assessment: Common standards for rigor and relevance in teaching academic subjects*. *Iowa Department of Education*. https://doi.org/10.1037/0022-0663.100.3.590

[30] Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144.

[31] Petchprasert, A. (2021). *Utilizing an automated tool analysis to evaluate EFL students' writing performances*. *Asian-Pacific Journal of Second and Foreign Language Education, 6*(1), Article 107. Springer Science+Business Media. https://doi.org/10.1186/s40862-020-00107-w

[32] Pui, P., Yuen, B., & Goh, H. (2020). *Using a criterion-referenced rubric to enhance student learning: A case study in a critical thinking and writing module*. *Higher*

Education Research & Development, 40(5), 1056. Taylor & Francis. https://doi.org/10.1080/07294360.2020.1795811

[33] Raman, A. (2014). TPACK Confidence of Pre-service Teachers in Universiti Utara Malaysia. *Mediterranean Journal of Social Sciences*. https://doi.org/10.5901/mjss.2014.v5n22p167

[34] Raman, A. (2019). Potentials of Fog computing in higher Education. *International Journal of Emerging Technologies in Learning (iJET)*, *14*(18), 194. https://doi.org/10.3991/ijet.v14i18.10765

[35] Raman, A., Ying, L. T., & Khalid, R. (2015). The Relationsip between Culture and Organizational Commitment among Chinese Primary School Teachers. *Mediterranean Journal of Social Sciences*. https://doi.org/10.5901/mjss.2015.v6n2s5p93

[36] Rathakrishnan, M., Raman, A., Haniffa, M. A., Mariamdaran, S. D., & Haron, A. (2018). The drill and practice application in teaching science for lower secondary students. *International Journal of Education, Psychology and Counseling*, *3*(7). http://repo.uum.edu.my/27063/

[37] Reynders, G., & Reddy, Y. M. (2021). *Developing and validating scoring rubrics for the assessment of research papers writing ability of EFL/ESL undergraduate students. Journal of Language Teaching and Research, 12(5),* 789-799. https://doi.org/10.17507/jltr.1205.10

[38] Riwayatiningsih, R. (2019). *Improving writing skill with questioning: A path on critical thinking skill. Metathesis Journal of English Language Literature and Teaching, 3*(2), 203. https://doi.org/10.31002/metathesis.v3i2.1665

[39] Setiawan, H., Khair, B. N., Ratnadi, R., Hakim, M., & Istiningsih, S. (2020). *Developing HOTS-based assessment instrument for primary schools. Proceedings of the 1st Annual Conference on Education and Social Sciences (ACCESS 2019).* https://doi.org/10.2991/assehr.k.200827.054

[40] Smith, A. J., Chen, L., & Rao, M. (2025). Rubrics and cognitive skill development: Addressing higher-order thinking in essay writing. *Contemporary Issues in Education, 15*(1), 68–82.

[41] Smith, A., & Johnson, B. (2025). *Engaging elementary students in higher order talk and writing about text. Journal of Early Childhood Literacy, 25(1),* 89-105. https://doi.org/10.1177/1468798417690918

[42] Sulaiman, F., et al. (2017). Effects of Inquiry-Based Approaches on Students' Higher-Order Thinking Skills. *Journal of Education and Practice*, 8(9), 135–145.

[43] Sulaiman, T., Subramaniam, P., & Kamarudin, N. (2019). *The influence of higher order thinking and metacognitive skills towards hands-on teaching among primary school science teachers. International Journal of Academic Research in Progressive Education and Development, 8*(4). https://doi.org/10.6007/ijarped/v8-i4/6497

[44] TESOL International Association. (2020). *Assessing writing. TESOL Blog.* https://doi.org/10.1093/applin/21.4.431

[45] Trilling, B., & Fadel, C. (2009). *21st Century Skills: Learning for Life in Our Times.*

[46] Zaki, M., Amalia, R., & Sofyan, S. (2020). *Development of high order thinking skills (HOTS) test instrument on exponent for junior high school students. Journal of Physics Conference Series, 1521*(3), 32096. IOP Publishing. https://doi.org/10.1088/1742-6596/1521/3/032096

[47] Zhang, Y. G., & Dang, M. Y. (2020). *Understanding essential factors in influencing technology-supported learning: A model toward blended learning success. Journal of*

Information Technology Education Research, 19, 489. Informing Science Institute. https://doi.org/10.28945/4597

[48] Zhang, Y., & Chen, L. (2024). *Enhancing English writing competence in higher education: Teacher-only assessment vs. combined teacher and student self-assessment*. *Smart Learning Environments, 11*, Article 263. https://doi.org/10.1186/s40862-024-00263-3

[49] ŽivkoviĿ, S. (2016). A Model of Critical Thinking as an Important Attribute for Success in the 21st Century. *Procedia - Social and Behavioral Sciences*, 232, 102–108.