Journal of Curriculum Studies Research

https://curriculumstudies.org

E-ISSN: 2690-2788

Volume: 7 Issue: 2 2025

pp. 155-176

Developing Students' HOTS Digital Literacy Profile by Using IRT

Guldana A. Begimbetova[a], Heri Retnawat[a], Gulzhaina K. Kassymova*[b], Mochamad Bruri Triyono[a], Mohamed Nor Azhari Azman[c] & Reisa Aulia Sodikin[d]

* Corresponding author
Email:g.kassymova@satbayev.university
a. Graduate School, Yogyakarta State University, Yogyakarta, Indonesia
b. Faculty of Pedagogy and Psychology, Abai Kazakh National Pedagogical University, Almaty, Kazakhstan; Institute of Metallurgy and Ore Beneficiation JSC, Satbayev University Almaty, Kazakhstan
c. Department Engineering Technology, Faculty of Technical and Vocational, Sultan Idris Education University, Malaysia.
d. Education of System and Information Technology, Indonesia University of Education, Bandung, Indonesia.

Article Info
Received: February 17, 2025
Accepted: July 10, 2025
Published: August 10, 2025

10.46303/jcsr.2025.16

How to cite
Begimbetova, G.A., Retnawat, H., Kassymova, G. K., Triyono, M. B., & Sodikin, R. A. (2025). Developing Students' HOTS Digital Literacy Profile by Using IRT. *Journal of Curriculum Studies Research*, 7(2), 155-176. https://doi.org/10.46303/jcsr.2025.16

ABSTRACT

This study aimed to develop and validate a digital literacy higher-order thinking skills (HOTS) cognitive test for engineering students at Universitas Negeri Yogyakarta in Indonesia. Utilising a quantitative research design, the study involved 95 participants and employed expert judgement to assess content validity through the Aiken V Index, which yielded a high average validity score of 1.783. Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA) confirmed a single factor structure with strong fit indices (CFI = 0.98, TLI = 0.97, RMSEA = 0.08), indicating robust construct validity. Reliability analysis revealed a high internal consistency (α = 0.92). Item Response Theory (IRT) calibration demonstrated satisfactory model fit, with moderate discrimination and average difficulty indices for the test items. The findings affirm that the developed test is both reliable and valid, effectively capturing students' digital literacy HOTS profiles, thereby providing valuable insights for educational interventions aimed at enhancing digital literacy skills in engineering education.

KEYWORDS

Digital literacy; item characteristics; HOTS cognitive test; reliability; validity; students; education.

INTRODUCTION

Determining students' higher-order thinking skills (HOTS) profiles within a digital media ecosystem is a crucial area of research for digital literacy test developers or measurement experts (Handayani et al., 2019). With the rapid integration of digital technologies into educational settings, it is essential to understand how students engage with digital media and how this engagement influences their cognitive abilities. The COVID-19 pandemic has further highlighted the significance of digital platforms for remote learning, making it imperative to investigate students' HOTS profiles (C4, C5, C6) within this context (Setiawati & Setyarini, 2020; Begimbetova et al., 2024; Sheriyev et al., 2016; Handayani et al., 2019; Nogaibayeva et al., 2023). Digital platforms or social media applications can support learners in accessing, co-constructing, and evaluating the authenticity of information (Lee et al., 2024; Begimbetova et al., 2023). In order to enhance students' higher-order thinking skills (HOTS) and digital literacy, Item Response Theory (IRT) has emerged as a robust methodological tool. Setiawan et al. (2024) demonstrated the utility of IRT in analyzing Differential Element Functioning (DIF) across regional contexts in national examinations, highlighting disparities that inform fair assessment design. In addition, Nursultankyzy et al. (2024) examined how language learning influences students' motivation and volunteering, and considered that engagement and cognitive development are closely linked.

However, online learning environments and other digital tools have transformed how students access information and construct knowledge. Unlike traditional learning contexts, these ecosystems are characterised by dynamic, interactive, and multimodal content that requires advanced cognitive engagement. The complexity of navigating such environments demands that students not only consume information but also critically evaluate its authenticity and relevance amidst widespread misinformation (Ma et al., 2021). Therefore, assessing how students apply HOTS in this multifaceted digital landscape is crucial for educators and policymakers seeking to prepare learners to thrive in an information-rich, yet often ambiguous, digital world (Lee et al., 2024; Maharani et al., 2022).

The COVID-19 pandemic has accelerated the reliance on digital learning platforms, highlighting the urgency of effectively measuring digital literacy skills (Getenet et al., 2024; Magocha et al., 2025). Remote and hybrid learning environments have made it necessary to track how well students can independently manage, interpret, and critically assess digital information sources such as Google Forms, Instagram, TikTok, and YouTube. While these platforms offer valuable educational opportunities, they also pose challenges, including exposure to fake or misleading information that can negatively impact learning outcomes (Abdillah et al., 2023). This study acknowledges the growing need to equip students with digital literacy skills that support safe, critical, and productive engagement within these platforms, making it timely and relevant to educational measurement (Isnah et al., 2022).

Methodologically, the study uses advanced psychometric techniques, including Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), and Item Response

Theory (IRT), to ensure that the developed test is both valid and reliable (Mardiana, 2024). This rigorous validation not only confirms the construct of digital literacy HOTS but also enables for precise analysis of item characteristics, such as difficulty and discrimination (Herunata & Aini, 2024). Understanding these parameters facilitates more targeted educational interventions and finer-grained insight into students' abilities, moving beyond traditional assessment methods to address the complexity of digital learning environments (Slamet, 2024). This methodological rigour enhances the test's capacity to meaningfully profile students' strengths and weaknesses.

The primary goal of this research is to develop a scientifically sound and practically applicable instrument for measuring students' HOTS within the digital media ecosystem. In doing so, this study contributes essential knowledge about students' digital literacy profiles, highlighting their capacity to engage critically with digital content. The findings will inform educators, curriculum developers, and policymakers in designing targeted strategies that foster higher-order cognitive skills necessary for digital citizenship and lifelong learning. The significance of this research is amplified by the ongoing digital transformation in education, where equipping students with strong HOTS-related digital literacy is fundamental to academic success and informed participation in society.

The study aims to develop a valid and reliable digital literacy cognitive test to measure higher-order thinking skills (HOTS) among engineering students. The test focuses on students ' ability to analyse, evaluate, and create digital content, thereby contributing to a broader field of educational assessment. Based on the background presented above, the authors formulate following research questions:

1. What are the specific characteristics of students' higher-order thinking skills (HOTS) in digital literacy as measured by the developed cognitive test, and how do these characteristics vary across different levels of digital media engagement?
2. How do the item parameters (difficulty, discrimination, and measurement precision), as identified through Item Response Theory (IRT), contribute to the understanding of students' digital literacy profiles in a rapidly evolving digital media ecosystem?

LITERATURE REVIEW

Digital literacy for higher-order thinking skills (HOTS) emphasizes the integration of critical thinking processes within the practical application of digital skills for problem-solving. It encompasses the ability to critically analyze digital data and information for accuracy, correctness, and validity (Badalov et al., 2020; Siritheeratharadol et al., 2023). It also involves evaluating digital content, comparing data, and interpreting information to make informed decisions (Gutiérrez-Porlán et al., 2022; Mutarah et al., 2024). Furthermore, digital literacy in the context of HOTS includes creating digital content by collecting and integrating digital data, as well as identifying and addressing digital skill gaps (Hamakali & Josua, 2023; Kassymova et al., 2019; Szyszka et al., 2022).

Theoretical perspectives on the development of higher- order thinking skills (HOTS)within the digital media ecosystem have gained increasing attention in recent years (Abosalem, 2015; Japar et al., 2025; Setyarini et al., 2023). Media ecology, as a metaphorical approach, offers opportunities to expand theoretical understanding and explore the dynamics of HOTS within the digital landscape (Insani et al., 2024). Additionally, the integration of critical theory and ecological perspectives has contributed to the formulation of new frameworks for HOTS development. The theoretical perspective of social ecology provides a comprehensive framework for examining the relationship between digital literacy and HOTS, enabling the translation of theoretical principles into practical guidelines for promoting HOTS in the digital realm (Valverde-Berrocoso et al., 2021).

Table 1.

HOTS in Digital Literacy Skills. Source: Churches

Digital literacy skills	**Descriptors**	**Cognitive (HOTS) levels**
Information and data literacy	Analysing digital data accuracy	Analysing C4
	Analysing digital information correctness	Analysing C4
	Analysing digital content validity	Analysing C4
	Comparing digital data	Evaluating C5
	Interpreting digital information	Evaluating C5
	Rating digital content	Evaluating C5
Digital content creation	Collecting digital data	Creating C6
	Creating digital content	Creating C6
	Integrating and re-elaborating / reconstructing digital information	Creating C6
Problem-solving	Identifying digital needs	Analysing C4
	Developing digital solutions to cater to the digital needs identified	Creating C6
	Creatively using digital technologies	Creating C6
	Identifying digital skill gaps	Analysing C4

In the digital media ecosystem, tasks and challenges often require the mobilization of digital literacy and the application of HOTS at the C4, C5, and C6 cognitive levels. Analyzing the accuracy and correctness of digital data, evaluating digital content, and comparing digital data are key components of HOTS at the C4 and C5 levels (Hafiza Hamzah et al., 2021). At the C6 level, individuals engage in creating digital solutions, planning digital projects, and creatively using digital technologies These tasks, which involve critical thinking, problem-solving, and information processing, demonstrate the practical use of digital literacy skills and the application of HOTS in the digital media ecosystem (Setyadi et al., 2025).

Although there is a particular emphasis on digital literacy and HOTS, existing grading systems often do not align with the actual cognitive descriptors and levels, particularly those outlined in Bloom's Digital Taxonomy. While research frequently explores digital skills, few studies explicitly examine their relationship to HOTS, and even fewer provide specific descriptors for tasks that reflect the cognitive demands at the C4 to C6 levels (Suhardiman et al., 2024). Even there is a substantial body of related literature, only a limited number of studies clearly present the descriptors and cognitive levels associated with HOTS. Based on the UNESCO digital framework, Table 1 provides an example of HOTS.

Tasks requiring the mobilization of digital literacy, particularly at the C4, C5, and C6 cognitive levels, involve various abilities. At the C4 level (Analysing), individuals engage in activities such as distinguishing and organising digital data, as well as attributing information sources (Aitken et al., 2011). The C5 level (Evaluating), involves checking and assessing digital data, as well as critically analysing digital content. Additionally, at the C6 level (Creating), individuals formulate digital solutions.

To develop accurate digital literacy tasks or assessments that foster HOTS, educators and test developers must possess a robust understanding of the digital literacy taxonomy, particularly the descriptors for cognitive levels C4, C5, and C6. These cognitive levels provide a framework for assessing students' abilities to engage in critical thinking and problem-solving within the digital realm.

Digital abilities mobilized at cognitive levels 4, 5, and 6 in Bloom's Digital Taxonomy. At cognitive level 4, students should demonstrate the ability to assess the accuracy of digital data, verify information correctness, and evaluate content validity. This involves evaluating the reliability and credibility of digital information sources, identifying potential biases, and critically examining data accuracy. At cognitive level 5, students should exhibit proficiency in assessing digital data by comparing multiple sources, interpreting digital information within context, and evaluating the quality and relevance of digital content. This level emphasizes the importance of students' capacity to make informed judgments about digital information and distinguish reliable sources from unreliable ones. At cognitive level 6, students should engage in the creation of digital content, which includes collecting digital data, integrating and re-elaborating digital information, and utilizing digital technologies creatively to address identified digital needs. This level focuses on students' ability to apply their knowledge and skills to generate original digital content and solutions. See Fig. 1.

Based on the HOTS digital literacy skills outlined in the table, a variety of digital literacy items can be developed. For instance, items can be designed to assess:

- students' proficiency in analysing digital data accuracy, such as providing a dataset and asking students to identify any inconsistencies or errors within it.
- learners' capacities to evaluate or assess the credibility and reliability of sources.
- students' ability to collect digital data, synthesise it, and present their findings in a multimedia format (e.g. video, podcast, posters, or infographics).

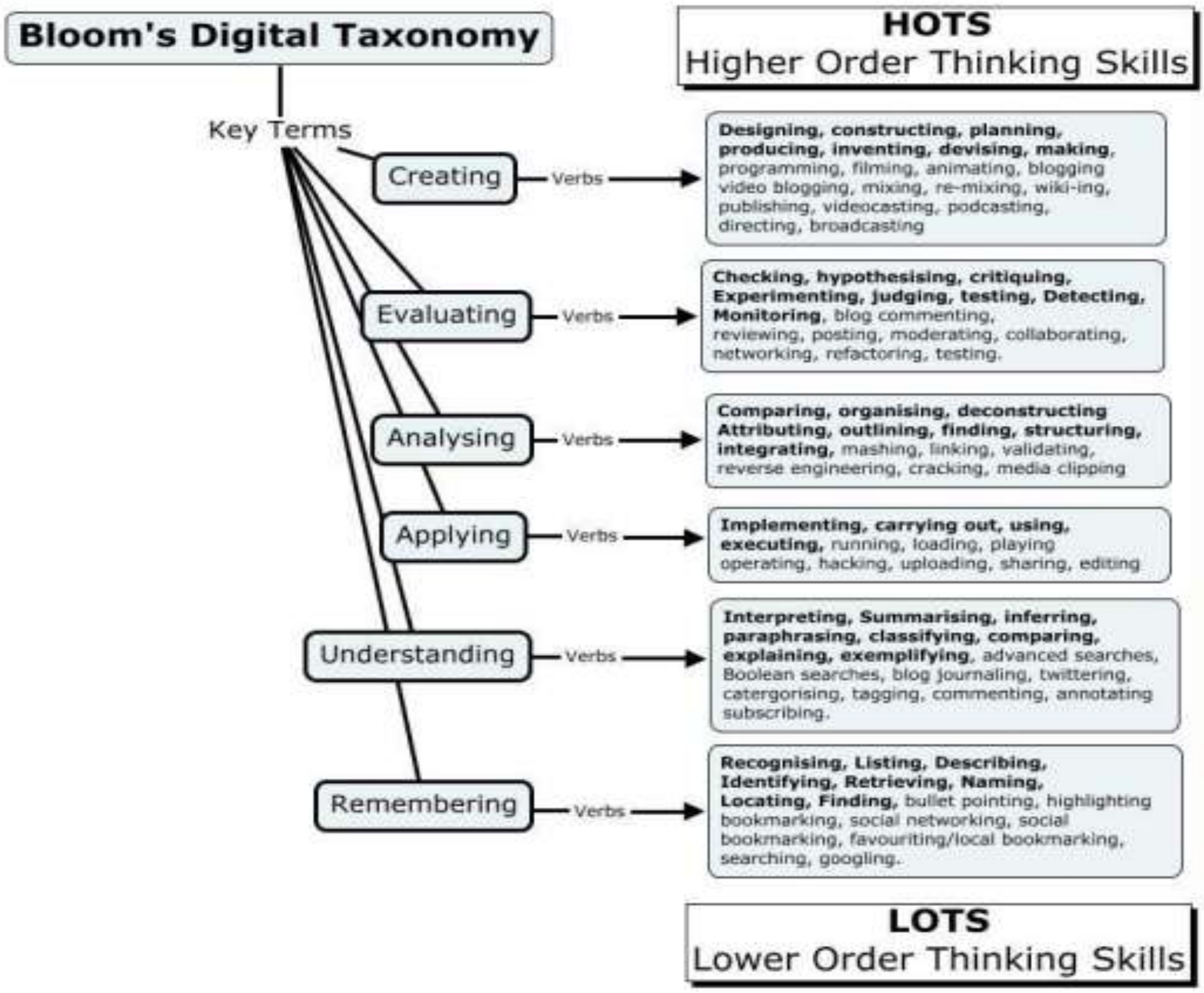
- students' skills in identifying digital needs, developing innovative digital solutions to address them, and showcasing their creative use of digital technologies.

Since it is crucial to investigate and determine students' HOTS within the digital media ecosystem, this research aims to:

- develop a reliable and valid digital literacy assessment for the HOTS cognitive test.
- calibrate the test items' parameter characteristics.
- determine students' HOTS in digital literacy profiles.

Figure. 1.

The Bloom's Digital Taxonomy for evaluating digital tasks revision



Although many studies have identified digital literacy, there is no consensus on its components, especially when combined with HOTS structures. For example, UNESCO emphasises the technical and critical use of digital media, while other studies prioritize ethical criteria and content creation. These inconsistencies cause ambiguity in the development and verification of the test. The study addresses this gap by aligning digital literacy with Bloom's

Digital Taxonomy levels (C4 to C6), providing the possibility of practical application in the context of theoretical compatibility and evaluation.

METHOD

Research Design

This study employed a quantitative research design, specifically used a psychometric approach to test development and validation to investigate the digital literacy skills of engineering students at Universitas Negeri Yogyakarta. A random sampling method was utilized to ensure the representativeness of the participants, allowing for generalizable findings regarding their discipline-specific digital literacy competencies. The research aimed to develop and validate a newly constructed assessment instrument that evaluates students' abilities to effectively utilize digital tools and resources within an English-language context, specifically tailored to the engineering discipline. This comprehensive assessment aimed to capture their digital literacy proficiency relevant to engineering. The methodology followed established practices in quantitative educational measurement by using statistical techniques to evaluate the validity and reliability of the assessment instrument. In this research implemented a multi-step quantitative validation process: Aiken V Index to assess content validity through expert judgment(Aiken, 1980),Exploratory Factor Analysis (EFA) to explore the underlying factor structure of the instrument, Confirmatory Factor Analysis (CFA) to confirm the factor structure and assess the model fit using indices such as CFI, TLI, RMSEA, and SRMR(Brown, 2015), Item Response Theory (IRT), specifically the Generalized Partial Credit Model (GPCM), to calibrate item parameters (difficulty and discrimination) and evaluate measurement precision. The study's methodological approach aligns with recommendations by psychometric researchers for rigorous test development in educational contexts (Begimbetova et al., 2022; Hambleton & Slater, 1997b; Kosherbayeva et al., 2024).

Participants

The construct validity of the digital literacy measure was established through a sequential validation approach. The sample consisted of 95 undergraduate engineering students. According to guidelines for factor analysis, a sample size of at least 5 to 10 participants per item is recommended (Osborne & Costello, 2004). Since the test included 7 items, the sample size meets the minimum threshold for both Exploratory Factor Analysis (EFA) and Confirmatory Factor Analysis (CFA). EFA was conducted to delineate the underlying factor structure, while CFA was used to confirm the identified structure and evaluate the model's fit indices. This rigorous validation process ensured the instrument's reliability and clarified the operationalization of the digital literacy construct. See Table 2.

Quantitative analytical methodologies were applied to scrutinize the data. Descriptive statistical measures were utilized to encapsulate the performance metrics of the participants on the assessment tool. The entire of the data analysis operations was facilitated by the R Studio software (version 2023.3.1.446), developed by the Posit team. The research incorporated a suite of complementary R packages. For executing the EFA, the ‘psych’ package was employed. The reliability and calibration, predicated on Classical Test Theory (CTT) were ascertained using the CTT package. Subsequently, for the calibration of items grounded in Item Response Theory (IRT), the ‘mirt’ package was utilized.

Table 2.

Participant demographics (n = 95)

Aspects	Mean
Gender	Male 5, Female 90
Year in program	1st-year
Academic level	S1 – Undergraduate
Study program	Engineering

Instrumentation

A newly developed instrument was employed to assess the digital literacy competencies of the participants. This instrument was designed to evaluate the students' proficiency in utilizing digital tools and resources relevant to their engineering studies. The construct validity of the digital literacy measure was established through a sequential validation approach, which included EFA to delineate the underlying factor structure. CFA was subsequently conducted to corroborate the identified structure and evaluate the model's fit indices, ensuring the reliability and clarity of the digital literacy construct being operationalized.

Data Collection Procedure

Data collection was conducted using the newly developed assessment instrument, administered to participants in a controlled environment. The assessment aimed to capture participants' engineering-relevant digital literacy proficiency, focusing on their ability to navigate and utilize various digital platforms and tools effectively. Participants were informed about the purpose of the study and provided consent prior to their participation.

Data Analysis

Quantitative analytical methodologies were employed to scrutinize the collected data. Descriptive statistical measures were utilized to summarize participants' performance metrics on the assessment tool. The entirety of the data analysis operations was facilitated by R Studio software (version 2023.3.1.446), developed by the Posit team. A suite of complementary R packages was incorporated into the analysis. The ‘psych’ package was employed for executing the EFA, while the reliability and calibration based on CTT were ascertained using the CTT package. For the calibration of items grounded in IRT, the ‘mirt’ package was utilized, ensuring a comprehensive analysis of the instrument's psychometric properties.

RESULTS

Content Validity

The Aiken V Index

During the trial stage, one of the validation types was content validity through expert judgment. The experts' feedback on the items developed became data used to compute the qualitative input on each item being assessed. R Studio software output for Aiken Index values was as follows:

Table 3.

Content Validity Scores

Item	Aiken V	Interpretation
X1	1.75	High validity
X2	1.70	High validity
X3	1.90	High validity

The results of the Aiken index calculations indicate strong content validity for the test under consideration. Individual rater scores, including Expert 1 (V = 1.75), Expert 2 (V = 1.70), and Expert 3 (V = 1.90), all surpassed the cut-off score for high validity, demonstrating a high level of agreement and consistency among the raters. Furthermore, the average Aiken index (V) across all raters was found to be 1.783333, which is above the threshold for high validity.

This average value, in conjunction with the individual rater scores, provides robust evidence of the test's content validity. Consequently, the overall categorization of 'high validity' further strengthens the argument that the test demonstrates strong content validity. The high level of agreement and consistency among the expert raters suggests that the test effectively measures the intended content domain, thereby contributing to its credibility and reliability in assessing the targeted construct.

Construct Validity Exploratory Factor Analysis

The evidence of fitness traceable in the R Studio software output related to Kaiser-Meyer-Olkin (KMO) factor adequacy and the MSA for each item is as follows:

Kaiser-Meyer-Olkin factor adequacy

Table 4.

Kaiser-Meyer-Olkin (KMO) Measures for Each Item

Item	KMO Value
X1	0.91
X2	0.93
X3	0.91
X4	0.92
X5	0.91
X6	0.89
X7	0.87
Overall MSA	0.90

The KMO values in the output indicate that the data is highly suitable for factor analysis. The overall KMO value is 0.90, which is well above the acceptable threshold. Additionally, the individual KMO values for each item (X1, X2, X3, X4, X5, X6, X7) are also above the acceptable level (Fig. 2), ranging from 0.87 to 0.93 (Fig. 3). These values suggest that the observed variables exhibit strong intercorrelations and provide meaningful information for extracting underlying factors through factor analysis.

Figure 2.

Scree plot eigen values

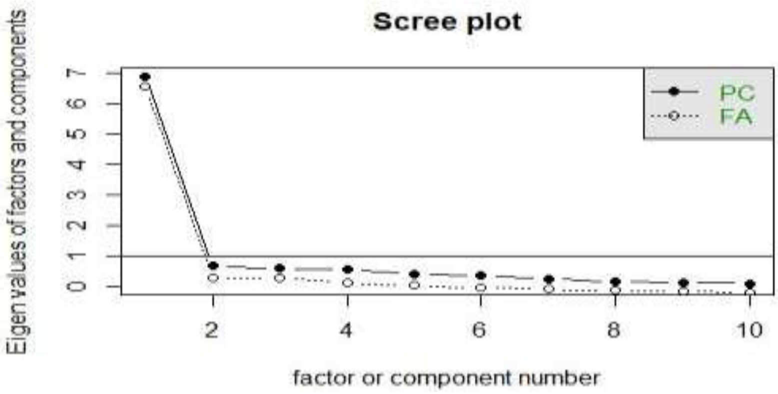
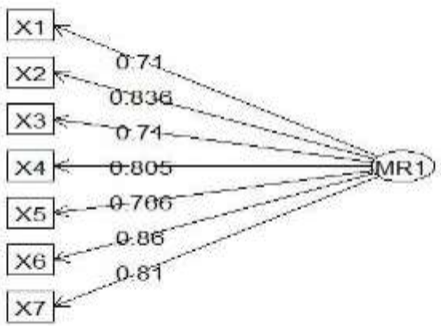


Figure 3.

Factor analysis



PC stands for Principal Components. Principal Component Analysis (PCA) is a statistical technique used to identify patterns and highlight variation within a dataset. It achieves this by converting potentially correlated variables into several uncorrelated variables called principal components.

FA, or Factor Analysis, is similar to PCA in that it reduces the dimensionality of data. However, unlike PCA, FA assumes that the variation in the data arises from underlying latent factors. The objective of this study is to identify the causal relationships among observed variables by determining the factors that contribute to their correlations. The KMO values in the output indicate that the data is highly suitable for factor analysis, as both the overall KMO value and the individual item KMO values exceed the acceptable threshold. These results suggest that the observed variables exhibit strong intercorrelations and offer valuable information for extracting underlying factors through factor analysis.

The goodness-of-fit index is 0.90, indicating a high level of agreement between the observed and predicted covariance values. A value close to 1 suggests a good model fit. The indicators X1 to X7 are closely associated with the construct or factor representing the digital literacy skills being assessed. The factor analysis plot from the EFA revealed standardised loadings, also known as the pattern matrix, which illustrate the relationships between the observed variables (X1 to X7) and the underlying latent factor (MR1). All indicators were retained in the measurement model, and the analysis yielded good fit statistics. This indicates a high level of agreement between the observed and predicted covariance values, suggesting that the indicators (X1 to X7) are closely associated with the tested construct or factor, validating the underlying theoretical framework of the test. See Fig. 3.

The standardised loadings in the factor analysis plot reveal the relationships between the observed variables (X1 to X7) and the underlying latent factor (MR1). These loadings, also known as the pattern matrix, indicate the strength and direction of the relationships between the observed variables and the underlying factor. In this case, all the indicators (X1to X7) show significant relationships with the latent factor, as indicated by the standardized loadings. The plot demonstrates that the observed variables are closely associated with the underlying construct of digital literacy skills being tested. The criterion of a 0.4 cut-off value is used, and all the indicators meet this criterion, indicating that they can be included in the measurement model. This suggests that the observed variables effectively capture the targeted construct, contributing to the credibility and reliability of the assessment.

Confirmatory Factor Analysis

Evidence supporting the measurement model requires the assessment of model fit using established cut-off criteria. To that end, the researchers considered the CFI, TLI > 0.90, and RMSEA, SRMR<0.08 as cut-off values. Based on these criteria, the CFA indicates a good model fit, as reflected in the values presented in Table 5.

Table 5.

CFA Statistics

No	Index	Value
1	CFI	0.98
2	TLI	0.97
3	RMSEA	0.08
4	SRMR	0.03
5	AIC	1221.67
6	BIC	1257.43

The TLI (Tucker-Lewis Index) and RMSEA (Root Mean Square Error of Approximation) indices provide evidence about the Good Fit of the CFA Model used. A TLI value of 0.98 and 0.97, along with RMSEA and SRMR indices of 0.08 and 0.03, respectively, indicate a strong fit for the CFA Model. These values meet the criteria for a good fit, as TLI and RMSEA values above 0.90

and below 0.08 respectively, are considered indicative of a well-fitting model. Therefore, the TLI and RMSEA indices provide strong evidence that the CFA model used has a good fit.

Reliability Index

The reliability of the digital literacy skills test, based on the output from R Studio software, is very high. The standardised alpha coefficient, which is a measure of internal consistency, is 0.92. This value indicates that the test is highly reliable in terms of its consistency:

Table 6.

Reliability Analysis of the Digital Literacy HOTS Test

Coefficient Type	Value
Raw Alpha	0.92
Standardized Alpha	0.92

Reliability Index to provide evidence that the test given was reliable, below is an output from R: Reliability analysis Call: alpha (x = data) raw_ alpha std. alpha 0.92 0.92 Taking into consideration the std. alpha, that is, Standardised Alpha, that value is 0.92. This means that it is higher than 0.80 and it is in the very high-reliability category. In summary, the digital literacy skills test is reliable in terms of its consistency.

IRT Item Parameter Calibration

The suitability of an IRT model is determined based on several criteria, including the p-value, RMSEA, CFI (Comparative Fit Index), TLI (Tucker-Lewis Index), AIC (Akaike Information Criterion), BIC (Bayesian Information Criterion), and SABIC (Sample-Size Adjusted BIC).

The determination of model fit indicates that each utilized IRT model has a p-value > 0.01, RMSEA < 0.8, CFI > 0.9, and TLI > 0.9. If these values are satisfied, the model is considered suitable for use. Additionally, to assess the appropriate model fit, one can compare the values of AIC, BIC, and SABIC across models. The model with the lowest AIC and BIC values indicates the best fit and is considered suitable for use.

The best model is chosen based on recommendations from Maydeu-Olivares and Paek & Cole (2019), using the smallest RMSEA and M2 values. Table 7 shows that the GRM and GPCM models have the smallest M2 and RMSEA values. However, in this study, the GPCM model is chosen if we consider its related smallest AIC, BIC, and SABIC values. Therefore, the GPCM model is selected as the most suitable IRT model.

Table 7.

IRT Model Fit

Model	M2	df	*p-value*	RMSEA	CFI	AIC	BIC	SABIC
PCM	10.34614	7	0.1697916	0.0713114	0.951668	1190.1	1211.7	1177.44
GPCM	1.998958	1	0.1574074	0.103088	0.898998	1182.1	1209.9	1165.84
GRM	0.000727	1	0.9784858	0	1.101055	1184.6	1253.6	1168.39

However, in terms of the number of valid items under each model, the application of all three IRT models resulted in six valid items, with Item 1 not fitting in any of the models (see Table 8). For this model, the following observations were made:

Table 8.

Model Returning in many Valid Items

No	Item	Chi-Square	p-value	Decision
1	X1	20.84	0.00	Not fit
2	X2	3.40	0.85	Fit
3	X3	5.28	0.51	Fit
4	X4	7.17	0.21	Fit
5	X5	12.21	0.06	Fit
6	X6	11.03	0.14	Fit
7	X7	6.29	0.51	Fit

According to the Chi-square test, item X1 has a Chi-square value of 20.84 and a p-value of less than 0.05 (denoted as 0.00). This indicates that it does not fit the model well. Items X2, X3, X4, X5, X6 and X7 have chi-square values ranging from 5.28 to 11.03 and p-values greater than 0.05 (ranging from 0.15 to 0.51), indicating that they fit the model well. This indicates that the items fit the model well and there is no statistically significant deviation from the expected model.

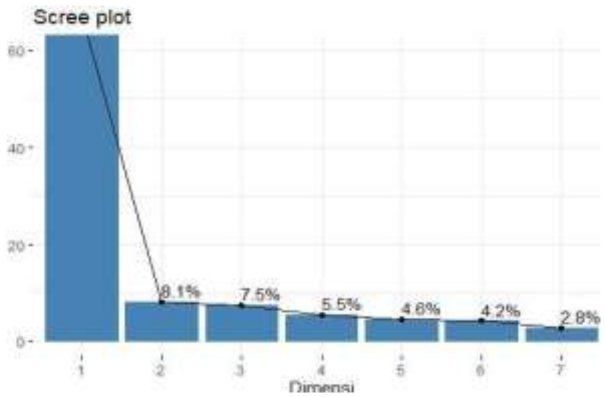
The decision criterion appears to be based on a common significance level (typically 0.05), with p-values below this threshold leading to a 'not fit' decision and p-values above this threshold leading to a 'fit"' decision. It is important to note that the chi-square test is useful for assessing model fit. However, it is also sensitive to sample size, and larger samples can lead to statistically significant results even when the model fit is acceptable. Therefore, it is important to interpret p-values in the context of the sample size of the study and other indices of model fit.

Unidimensionality Evidence

There is strong evidence supporting the test's unidimensionality, as illustrated in Figure 4. The scree plot can be seen in Fig. 4. The plot indicates that the instrument has a single dimension, visually represented by a steep slope. Multiple slopes would indicate multiple dimensions (Retnawati et al., 2016).

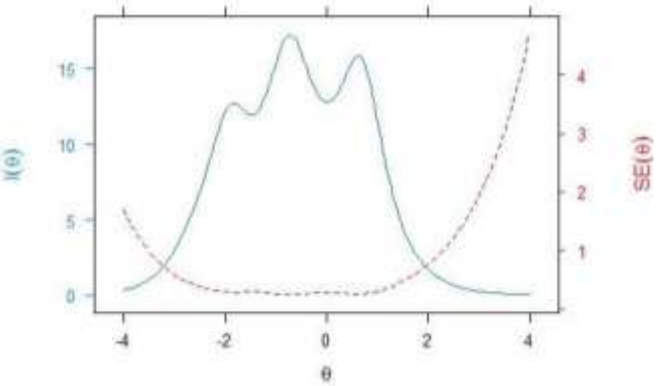
However, the ratio of the first eigenvalue to the total eigenvalue is > 25% (>60.0% in this case). This result indicates that the first factor contributes more than 20% of the variance (Hambleton & Slater, 1997a).

Figure 4.
Test Unidimensionality



Information Function and Standard Error
The information function of the SAIU instrument illustrates the extent to which a test can be effectively used for a specific ability. The shaded area in Fig. 5 indicates the intersection of the information function and the standard error.

In Fig. 5, it can be observed that the coordinate intersection point of the information function and standard error is at (-3) and (2). The abscissa coordinate of this intersection point explains the range of abilities suitable for taking this test or the suitability of using this test instrument for students with abilities ranging from less than -3 (low) to 2 (moderate). So, by utilizing the Information Function and Standard Error test, users of the SAIU instrument can gain insights into its effectiveness in measuring abilities and analyze the strengths and weaknesses of each item within the instrument.
Figure 5.
Test's SAIU



Items parameters characteristics
Discriminant Coefficient (D):
[1] 0.2302953
Categorization of Discriminant Index:
[1] "Fair"
[1] 0.2736842

Category of Average Difficulty Index:
[1] "Average"

The summary table shows the item characteristics for each item in the digital literacy HOTS test. Each item is assigned a difficulty index (p_Index) and a difficulty category (p_Category). The difficulty index represents the proportion of students who answered the item correctly. The difficulty categories are labeled as 'Difficult,' 'Average,' or Easy' based on the cutoff values provided.

The discriminant coefficient (D_Coefficient) is a measure of item discrimination. It quantifies how well an item can differentiate between high and low digital literacy HOTS test-takers. In this case, the discriminant coefficient is 0.2302953, indicating moderate discrimination.

On the one hand, the categorization of the discriminant index places it in the 'Fair' category, suggesting that the items have some discriminatory power but could be improved for better differentiation. The average difficulty index is 0.2736842, placing it in the 'Average' category. This indicates that, on average, the items are neither too difficult nor too easy for the students. To sum up, the item characteristics and indices suggest that the digital literacy HOTS test has moderate discrimination and falls within an acceptable range of difficulty, making it a suitable instrument for assessing the intended construct. Refer to Table 9.

Table 9.

Item characteristics

Item	p_index	p_category	D_coefficient	D_category
1	0.3789474	Average	0.2302953	Fair
2	0.2631579	Average	0.2302953	Fair
3	0.1578947	Difficult	0.2302953	Fair
4	0.2421053	Average	0.2302953	Fair
5	0.1473684	Difficult	0.2302953	Fair
6	0.2631579	Average	0.2302953	Fair
7	0.4631579	Average	0.2302953	Fair

Moderate item discrimination indicates that the test can generally differentiate between students with different HOT levels. However, to enhance diagnostic precision, it's essential to refine the items carefully. For instance, elements with high discriminatory values can effectively identify students who require targeted support. The inconsistencies observed in item X1 suggest that teachers revise test tasks not only in terms of content, but also in terms of clarity, compliance with cognitive level (C4 to C6), and cultural/contextual compliance. To foster higher cognitive engagement, particular attention should be given to elements targeting the **c**reating level (C6).

DISCUSSION

The findings of this research affirm the psychometric validity devoloped digital literacy HOTS cognitive test. The high content validity confirmed through the Aiken's V index (avereage V= 1.78) aligns that whre a V value above 0.80 indicates strong agreement among the experts (Aiken,1980). This result is consistent with guidelines from previous research (Sweileh, 2024) reinforces the notion that the digital literacy HOTS test effectively measures the intended construct. The EFA further strengthens these findings by supporting a single-factor structure with a high KMO value of 0.90, exceeding the min criteion of 0.60 suggested by (Kaiser, 1974), the item KMO values ranging from 0.87 to 0.93. These values indicate is very suitable for factor analysis and the items are sufficient variance.This is consistent with the KMO value of ≥0.80 for psychometric fit in scale construction, as recommended by Osborne and Costello (2004).These findings contribute to the understanding of the underlying structure of the digital literacy HOTS test and its capacity to capture the targeted construct accurately.

The Confirmatory Factor Analysis (CFA) conducted in this study demonstrates a strong fit between the proposed measurement model and the observed data, supported by the high values of TLI (0.98, 0.97) and the acceptable levels of RMSEA (0.08) and SRMR (0.03) all of which fall withn the recommende thresholds(Hu & Bentler, 1999), indicating the aimed measurement model fits the observed data well.

These metrics are similar to those that Brown (2015) described using to validate cogniti ve tests in online learning environments.These results provide robust evidence for the suitability of the identified factor structure and the evaluated measurement model. Additionally, the reliability analysis yielded a high level of internal consistency, with a standardised alpha coefficient of 0.92, conforming that the test items consistently measure the same underlying construct. This degree of reliability is comparable to that of Gutiérrez-Porlán et al. (2022), who used comparable constructs to assess digital abilities and reported alpha values as high as 0.90. This strengthens the digital literacy HOTS test's stability and dependability.

The generalized partial credit model (GPCM) in conjunction with item response theory (IRT) enabled thorough item-level calibration. The chosen model demonstrated acceptable limits for TLI (> 0.90), RMSEA (< 0.08), CFI (> 0.90), and p-value (> 0.01), indicating that the model fit data well (Maydeu-Olivares & Joe, 2014). The test has a reasonable ability to distinguish between pupils who perform well and those who do not, according to the item discrimination index (D = 0.23), which is classified as "fair." This is acceptable for early instrument development, even though it is not high, as shown by Hambleton and Slater (1997).

Furthermore, the unidimensionality assumption was supported by the screen plot and eigenvalue analysis, with more than 60% of the total variance explained by the first factor, which is generally higher than the 20% benchmark by (Retnawati, 2016)This test provides additional evidence that HOTS measures a coherent and singular latent construct digital literacy.

These results broaden the theoretical and empirical foundation for evaluating HOTS in digital contexts, especially at the cognitive levels C4 (analyzing), C5 (evaluating), and C6 (creating) as described in Bloom's Taxonomy of Digital Literacy, when compared to related studies like Setyarini et al. (2023). This instrument distinguishes itself by integrating critical thinking, evaluation, and creative problem solving—dimensions that are crucial for developing digital citizenship and 21st century competencies (Valverde-Berrocoso et al., 2021), whereas many prior assessments have concentrated on technical or procedural digital skills e.g., Szyszka et al. (2022).

The research question of developing a reliable and valid digital literacy HOTS cognitive test has been answered through the discussion of the findings. The study provides evidence of high content validity, supported by established guidelines. The EFA and CFA demonstrate a good fit of the measurement model, indicating the test's reliability and construct validity. These findings confirm that the digital literacy HOTS cognitive test developed in this study is both reliable and valid.

The research question of determining students' HOTS digital literacy profiles has been answered based on the discussion of the results. Analysis of item characteristics, including discrimination and difficulty indices, provides valuable insights into students' performance and proficiency levels. The categorization of these indices allows for the identification of diverse student profiles regarding HOTS digital literacy skills. Therefore, the research question related to determining students' HOTS digital literacy profiles has been addressed and answered through the analysis of the results.

CONCLUSION

This study successfully developed and validated a digital literacy HOTS cognitive test tailored for engineering students at Universitas Negeri Yogyakarta. The findings indicate that the test possesses high content validity, as supported by established guidelines from previous research. The rigorous validation process, which included EFA and CFA, confirmed a single-factor structure that accurately reflects the underlying construct of digital literacy HOTS. This strong alignment between the observed data and the proposed measurement model underscores the test's effectiveness in capturing the intended cognitive skills.

Moreover, the reliability analysis revealed a high level of internal consistency, with a standardized alpha coefficient of 0.92, affirming that the test items consistently measure the same construct. The application of IRT further enhanced the robustness of the findings, as the calibration of item parameters met the established model fit criteria. This comprehensive approach to validation not only addresses the research question regarding the reliability and validity of the digital literacy HOTS cognitive test but also establishes a solid foundation for its use in educational assessments. The study also successfully addressed the research question concerning the calibration of test items' parameter characteristics. The analysis demonstrated that the item parameters were effectively calibrated, providing valuable insights into the difficulty and discrimination indices of the test items. This calibration process is crucial for understanding how well the test differentiates between varying levels of student proficiency in digital literacy HOTS, thereby enhancing the test's utility in educational settings. Educational institutions should consider integrating validated digital literacy assessments into the curriculum as diagnostic tools. Such tools can serve as the basis for developing essential digital skills among students during their application, inform personalized training, and guide curriculum reform. Decision-makers in the field of education can utilize research findings to implement specific changes that meet students' needs. The results obtained will help to improve curricula and determine the steps based on the development of digital literacy.

REFERENCES

Abdillah, H. Z., Partino, & Madjid, A. (2023). Enhancing Student Well-being through AI Chat GPT in the Smart Education University Learning Environment: A Preliminary Review of Research Literature. *E3S Web of Conferences*, *440*, 1–7. https://doi.org/10.1051/e3sconf/202344005005

Abosalem, Y. (2015). Assessment techniques and students' higher-order thinking skills. *ICSIT 2018 - 9th International Conference on Society and Information Technologies, Proceedings*, *March*, 61–66. https://doi.org/10.11648/j.ijsedu.20160401.11

Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, *40*(4), 955–959. https://doi.org/10.1177/001316448004000419

Aitken, J. E., Fairley, J. P., & Carlson, J. K. (2011). Communication technology for students in special education and gifted programs. In *Communication Technology for Students in Special Education and Gifted Programs* (pp. 1–399). https://doi.org/10.4018/978-1-60960-878-1

Badalov, A.A., Brovkina, S.N., Arpentieva, M.R., Kalinin, S.S., Kassymova, G.K. (2020). The Archetype of Intellectual Activity: A Modern Methodology for the Description of the Protophenomenon. *Clinical Psychology and Special Education*, 9(1), 1–16. (In Russ.). https://doi.org/10.17759/cpse.2020090101

Begimbetova, G., Abdigapbarova, U., Abdulkarimova, G., Pristupa, E., Issabayeva, D., & Kurmangaliyeva, N. (2022). Use of ICT in CLIL-classes for the Future Teachers Training. *ACM International Conference Proceeding Series*, 98 – 104. https://doi.org/10.1145/3543407.3543424

Begimbetova, G.A., Retnawati, H., Ndayizeye, O., Flindt, N., Kassymova, G.K. (2024). A Bibliometric Review on Exploring Digital Literacy Assessment Dynamics in Education. Challenges of Science. Issue VII, pp. 26-37. https://doi.org/10.31643/2024.04

Begimbetova G.A., Retnawati H., Nogaibayeva A.A., Sansyzbayeva D.B., Triyono M.B. (2023). Bibliometric Analysis of Research Related to Digital Literacy Using the Scopus Database from 2017-2023. Challenges of Science. Issue VI, 2023, pp. 5-14. https://doi.org/10.31643/2023.01

Brown, T. A. (2015). Confirmatory factor analysis for applied research, 2nd ed. In *Confirmatory factor analysis for applied research, 2nd ed.* The Guilford Press.

Getenet, S., Cantle, R., Redmond, P., & Albion, P. (2024). Students' digital technology attitude, literacy and self-efficacy and their effect on online learning engagement. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-023-00437-y

Gutiérrez-Porlán, I., Prendes-Espinosa, P., & Sánchez-Vera, M. D. M. (2022). Digital Technologies for the Assessment of Oral English Skills. *Applied Sciences (Switzerland)*, *12*(22). https://doi.org/10.3390/app122211635

Hafiza Hamzah, N., Khalid, M., & Wahab, J. A. (2021). The effects of principals' digital leadership on teachers' digital teaching during the covid-19 pandemic in malaysia. *Journal of Education and E-Learning Research*, *8*(2), 216–221. https://doi.org/10.20448/journal.509.2021.82.216.221

Hamakali, H., & Josua, L. (2023). Engendering Technology-Assisted Pedagogy for Effective Instructional Strategy in the University of Namibia Language Centre. *Research in Educational Policy and Management, 5*(1), 18-32. https://doi.org/10.46303/repam.2023.3

Hambleton, R. K., & Slater, S. C. (1997a). Item response theory models and testing practices: Current international status and future directions. *European Journal of Psychological Assessment*, *13*(1), 21–28. https://doi.org/10.1027/1015-5759.13.1.21

Hambleton, R. K., & Slater, S. C. (1997b). Item Response Theory Models and Testing Practices: Current International Status and Future Directions ⋆ ⋆Invited paper presented at the 23rd International Congress of Applied Psychology, Madrid, Spain. ⋆Laboratory of Psychometric and Evaluative Research Report No.229. Amherst, MA: University of Massachusetts, School of Education. *European Journal of Psychological Assessment*, *13*(1), 21–28. https://doi.org/10.1027/1015-5759.13.1.21

Handayani, H., Sopandi, W., Syaodih, E., Suhendra, I., & Hermita, N. (2019). RADEC: An Alternative Learning of Higher Order Thinking Skills (HOTs) Students of Elementary School on Water Cycle. *Journal of Physics: Conference Series*, *1351*(1). https://doi.org/10.1088/1742-6596/1351/1/012074

Herunata, H., & Aini, Z. N. (2024). The Analysis of Higher Order Thinking Skills: Transfer Aspect of Student Concept in Reaction Rate Material. *E3S Web of Conferences*, *481*. https://doi.org/10.1051/e3sconf/202448104006

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. In *Structural Equation Modeling* (Vol. 6, Issue 1, pp. 1–55). Lawrence Erlbaum. https://doi.org/10.1080/10705519909540118

Insani, A. N. P., Aprilia, P. K., & Mistar, J. (2024). Challenges and Opportunities: A Qualitative Study of EFL Teachers' Digital Proficiency and its Impact on Interactive Language Assessment. *Voices of English Language Education Society*, *8*(1), 43–51. https://doi.org/10.29408/veles.v8i1.25095

Isnah, E. S., Suyatno, & Subandiyah, H. (2022). The Effect of Digital Literacy on Language Ability in Higher Education: Experience From a Developing Country. *Journal of Higher Education Theory and Practice*, *22*(11), 215–222. https://doi.org/10.33423/jhetp.v22i11.5425

Japar, M., Hermanto, H., Muyaroah,S., Susila, H.R. & Alfani, H. (2023). Digital Literacy-Based Multicultural Education through Civic Education in Indonesian Junior High Schools, *Journal of Social Studies Education Research, 14(*4), 328-349. https://jsser.org/index.php/jsser/article/view/5281/651

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, *39*(1), 31–36. https://doi.org/10.1007/BF02291575

Kassymova, G.K., Duisenbayeva, Sh.S., Adilbayeva, U.B., Khalenova, A.R., Kosherbayeva, A.N., Triyono, M.B., Sangilbayev, O.S. (2019). Cognitive competence based on the e-learning. *International Journal of Advanced Science and Technology*, Vol. 28, Issue 18, 167-177.

Kosherbayeva, A. N., Issaliyeva, S., Begimbetova, G. A., Kassymova, G. K., Kosherbayev, R., & Kalimoldayeva, A. K. (2024). An overview study on the educational psychological assessment by measuring students' stress levels. *Cakrawala Pendidikan*, *43*(1), 1 – 18. https://doi.org/10.21831/cp.v43i1.66276

Lee, H. Y., Chen, P. H., Wang, W. S., Huang, Y. M., & Wu, T. T. (2024). Empowering ChatGPT with guidance mechanism in blended learning: effect of self-regulated learning, higher-order thinking skills, and knowledge construction. *International Journal of Educational Technology in Higher Education*, *21*(1). https://doi.org/10.1186/s41239-024-00447-4

Ma, K., Chutiyami, M., Zhang, Y., & Nicoll, S. (2021). Online teaching self-efficacy during COVID-19: Changes, its associated factors and moderators. *Education and Information Technologies*, *26*(6), 6675–6697. https://doi.org/10.1007/s10639-021-10486-3

Magocha, M., Munyaradzi, J., & Babalola, S. S. (2025). The Impact of the Pandemic on Digital Literacy Skills for Online Teaching in Zimbabwean Schools: A Mixed-Methods Research Approach. *Research in Social Sciences and Technology, 10*(1), 310-331. https://doi.org/10.46303/ressat.2025.17

Maharani S.D., Susanti R., Indarti L.H., Syamsi A. (2022). Integrating HOTS-Based Student Electronic Worksheet: Teaching Styles in Elementary School During the COVID-19 Pandemic. *Journal of Social Studies Education Research, 13* (3), pp. 98 - 119

Mardiana, H. (2024). Perceived Impact of Lecturers' Digital Literacy Skills in Higher Education Institutions. *SAGE Open*, *14*(3), 1–11. https://doi.org/10.1177/21582440241256937

Mutarah, R., Azman, M.N.A., Kassymova, G.K., Kenzhaliyev, B.K. (2024). Android-Based Interactive Application Development in the Subject of Design and Technology for the Topic of Manufacturing Technology. *AIP Conf. Proc*. 2750, 040065. https://doi.org/10.1063/5.014927222

Nogaibayeva, A.A., Kassymova, G.K., Triyono, S., Winantaka, B. (2023). Language teachers' ICT up-take in a single university in the developing country Kazakhstan. *Cakrawala Pendidikan,* Vol. 42, Issue 2, pp. 295- 309. https://doi.org/10.21831/cp.v42i2.57488

Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. In *Multivariate Behavioral Research* (Vol. 49, Issue 4, pp. 305–328). Taylor & Francis. https://doi.org/10.1080/00273171.2014.911075

Nursultankyzy, A. M., Khan, N. N., Kassymova, G. K., Kalimoldayeva, A. K., & Ketebayeva, E. K. (2024). Effects of learning languages on volunteering. *REID (Research and Evaluation in Education)*, 10(2). https://doi.org/10.21831/reid.v10i2.72655

Osborne, J. W., & Costello, A. B. (2004). Sample size and subject to item ratio in principal components analysis. *Practical Assessment, Research and Evaluation*, *9*(11), 1–9.

Retnawati, H. (2016). Validitas reliabilitas & Karakteristik butir: panduan untuk peneliti, mahasiswa, dan psikometrian. *Parama Publishing*, 1–212.

Sheriyev, M.N., Atymtayeva, L.B., Beissembetov, I.K., Kenzhaliyev, B.K. (2016). Intelligence system for supporting human-computer interaction engineering processes. *Applied Mathematics and Information Sciences*, Volume 10, Issue 3, pp. 927-935. https://doi.org/10.18576/amis/100310

Setiawan, A., Kassymova, G. K., Mbazumutima, V., & Agustyani, A. D. (2024). Differential Item Functioning of the region-based national examinationequipment. *REID (Research and Evaluation in Education)*, 10(1). https://doi.org/10.21831/reid.v10i1.73270

Setiawati, D., & Setyarini, S. (2020). Promoting EFL young learners' higher order thinking skills (HOTs) through interactive digital storytelling. *ACM International Conference Proceeding Series*, 57–61. https://doi.org/10.1145/3436756.3437021

Setyadi, A., Pawirosumarto, S., Damaris, A., & Dharma, R. (2025). Risk management, digital technology literacy, and modern learning environments in enhancing learning innovation performance: A framework for higher education. *Education and Information Technologies*, *2023*(0123456789). https://doi.org/10.1007/s10639-025-13380-4

Setyarini, S., Salim, H., & Purnawarman, P. (2023). Higher-Order Thinking Skills (HOTS)-based literacy media: An innovative learning strategy to promote the secondary students' social awareness. *Forum for Linguistic Studies*, *5*(2), 1–19. https://doi.org/10.59400/FLS.V5I2.1706

Siritheeratharadol, P., Tuntivivat, S., Intarakamhang, U. (2023). European Journal of Educational Research. *European Journal of Educational Research*, *12*(2), 749–758.

Slamet, J. (2024). Potential of ChatGPT as a digital language learning assistant: EFL teachers' and students' perceptions. *Discover Artificial Intelligence*, *4*(1). https://doi.org/10.1007/s44163-024-00143-2

Suhardiman, S., Margana, Putro, N. H. P. S., Hakiki, M., & Fadli, R. (2024). Using the Buana Online Course web-based mobile application to improve English for specific purpose engineering courses. *International Journal of English Language and Literature Studies*, *13*(3), 449–463. https://doi.org/10.55493/5019.v13i3.5189

Szyszka, M., Tomczyk, Ł., & Kochanowicz, A. M. (2022). Digitalisation of Schools from the Perspective of Teachers' Opinions and Experiences: The Frequency of ICT Use in Education, Attitudes towards New Media, and Support from Management. In *Sustainability (Switzerland)* (Vol. 14, Issue 14). https://doi.org/10.3390/su14148339

Valverde-Berrocoso, J., Fernández-Sánchez, M. R., Dominguez, F. I. R., & Sosa-Díaz, M. J. (2021). The educational integration of digital technologies preCovid-19: Lessons for teacher education. *PLoS ONE*, *16*(8 August), 1–22. https://doi.org/10.1371/journal.pone.0256283