Evaluating the Implementation of Computer-Based Assessment in Seafarer Certification: The Role of Baseline Competency Management and CBA Utilization on Graduation Rates through Tryouts

Retno Sawitri Wulandari[a], Deo Prawira[a], Larsen Barasa[a], B. B Harianto[b]

[a] Sekolah Tinggi Ilmu Pelayaran, Jl. Marunda Makmur, Daerah Khusus Ibukota Jakarta 14150 Inonesia
[b] Politeknik Penerbangan Surabaya, Jl. Jemur Andayani I No.73, Surabaya, 60236 Indonesia
[1] shota_ku82@yahoo.com*; [2] deo.prawira@gmail.com; [3] larsenbarasa@gmail.com, [4]bambangfarzardy@gmail.com
* corresponding author

ARTICLE INFO

Article history
Received April 06, 2026
Revised May 15, 2026
Accepted June 19, 2026

Keywords
Computer-Based Assessment (CBA)
Maritime Education
Certification Success
Baseline Competency
PLS-SEM
Tryout Simulation
Graduation Rate

ABSTRACT

The implementation of Computer-Based Assessment (CBA) in the Seafarer Professional Examination (UKP Pelaut) is a strategic policy to improve integrity, standardization, and accountability in maritime certification. As part of digital transformation, CBA aims to strengthen exam quality and workforce readiness, yet pass rates remain inconsistent, highlighting gaps between regulatory goals and outcomes. This study evaluates CBA effectiveness through a mediated structural model linking utilization, competency management, tryout performance, and graduation rates. Using data from 156 Level I participants (ANT I and ATT I) at STIP Jakarta, analyzed with PLS-SEM, results show that CBA significantly improves tryout performance (β = 0.686, $p < 0.05$) and graduation outcomes (β = 0.324, $p < 0.05$). Baseline competency also strongly affects readiness (β = 0.388, $p < 0.05$) and graduation (T = 3.611, $p < 0.05$), with tryout performance partially mediating these effects. Findings indicate that CBA is partially effective: while digital assessment enhances efficiency and responsiveness, full success requires integrated competency management and structured tryout mechanisms to ensure sustainable improvement in maritime certification.

I. Introduction

The implementation of Computer-Based Assessment (CBA) within professional certification systems represents not only a technological transition but also a strategic public policy intervention aimed at improving assessment integrity, standardization, and accountability for outcomes. In the maritime sector, the adoption of Computer-Based Assessment for the Seafarer Professional Examination (UKP Pelaut) is grounded in regulatory mandates issued by the Directorate General of Sea Transportation under the Ministry of Transportation, aligning with international competency frameworks such as the Standards of Training, Certification, and Watchkeeping (STCW). These regulations formalize CBA as an official assessment mechanism intended to enhance objectivity, transparency, and alignment with global maritime competency standards. Thus, the implementation of CBA UKP constitutes a policy instrument designed to strengthen certification quality and workforce readiness.

The primary objective of this policy is to ensure that professional certification processes are valid, reliable, and aligned with standardized competency benchmarks. Computer-based assessment systems (CBAS) have become increasingly central to professional education, particularly in disciplines where certification outcomes determine workforce eligibility and professional mobility. Empirical research consistently indicates that CBAS positively influences graduation rates and certification outcomes by enhancing performance accuracy, offering immediate feedback, and reducing examination-related anxiety (Wilson, 2018; Kaspar et al., 2016; Guimar√£es et al., 2017). Formative computer-based testing, for example, has been shown to significantly improve licensure examination performance in diagnostic medical sonography, thereby strengthening alignment between academic preparation and labor-market demands (Wilson, 2018). In nursing education, CBAS implementation has enabled systematic monitoring of competency attainment and learning outcomes, supporting reliable graduation and qualification standards (Kaspar et al., 2016; Guimar√£es et al., 2017). These findings support the policy rationale that digital assessment enhances institutional accountability and the credibility of certification.

In competency-driven sectors, assessment systems are closely tied to standardized professional frameworks that define required skills and knowledge. Competency standards such as STCW exemplify how structured benchmarks guide assessment design and certification processes to ensure alignment with industry. Standardized competency frameworks inform the structure and content of testing, reinforcing consistency between educational outcomes and workplace expectations (Kaspar et al., 2016). Adherence to such frameworks enhances the validity, reliability, and professional relevance of assessments, thereby supporting certification success

Vol. 9, Issue 1, June 2026, pp. 01-15 e-ISSN: 2654-5667

(Grosseck et al., 2023; Ovtšarenko & Safiulina, 2025). Consequently, the integration of CBAS within competency-based systems is expected to optimize both examination integrity and professional preparedness, fulfilling the normative goals of the CBA UKP policy.

However, despite its regulatory foundation and normative objectives, implementation challenges persist. The transition toward digital examination systems poses substantial constraints, including technological infrastructure limitations, uneven digital literacy among faculty, and the risk of academic dishonesty (Butler-Henderson & Crawford, 2020; Grosseck et al., 2023). Reports of increased cheating during unsupervised online examinations have raised serious concerns regarding assessment integrity in digital contexts (Castro et al., 2023). Technical disruptions during assessments, including interface limitations and navigation constraints, may negatively affect student scores and completion times (Cochran et al., 2020). Furthermore, institutions must invest in systematic training and support to ensure that both instructors and students can effectively adapt to digital assessment environments (Costa & Chen, 2023; Ceka & O'Geen, 2019). These implementation barriers indicate a potential gap between policy design and operational outcomes.

Low pass rates in professional certification examinations further signal the need to evaluate whether the policy objectives of CBA implementation are being effectively realized. Certification outcomes are frequently influenced by multifaceted determinants, including examination anxiety (Kolagari et al., 2018), discrepancies between traditional and digital assessment familiarity (Tian et al., 2018; Guimar√£es et al., 2017), and broader contextual factors such as socioeconomic conditions and access to technological resources (Costa & Chen, 2023; Webb et al., 2017). These determinants suggest that technological adoption alone may not guarantee improved graduation rates, thereby raising critical evaluative questions regarding policy effectiveness.

Familiarity with computer-based testing (CBT) platforms has been identified as a key moderating factor influencing examination performance and confidence. Regular exposure to digital assessment tools is associated with reduced anxiety, enhanced self-efficacy, and improved academic outcomes (Nardi & Ranieri, 2018; Nagy & Korom, 2023; Kolagari et al., 2018). From a cognitive load perspective, familiarity reduces extraneous processing demands, enabling learners to focus more effectively on task-relevant cognitive activities (Wang et al., 2022). Structured tryouts or mock examinations function as preparatory instruments that operationalize this familiarity. Formative mock testing has demonstrated measurable benefits, including reduced test-related anxiety and improved performance onachiev licensure examinations (Wilson, 2018). Moreover, structured mock assessments provide actionable feedback that enables candidates to refine preparation strategies prior to final evaluations (Wang et al., 2022; Ponce et al., 2020). These mechanisms suggest that the effectiveness of CBA policy implementation may depend not only on technological availability but also on competency management and structured rehearsal systems.

Although prior studies substantiate the individual contributions of CBAS, competency standards, digital familiarity, and mock examinations, these elements are frequently examined in isolation. Limited research evaluates the implementation of computer-based assessment as a policy system, with success measured by outcome indicators such as graduation rates. There is insufficient empirical modeling of how CBA utilization and baseline competency management interact through tryout mechanisms to influence certification outcomes within sector-specific training systems governed by regulatory mandates.

Therefore, this study positions itself as an outcome-based evaluation of the implementation of the computer-based assessment policy in the Seafarer Professional Examination (UKP Pelaut). By developing and empirically testing a mediated structural model linking CBA utilization, baseline competency and knowledge, tryout performance, and graduation rates, this research evaluates whether the policy's operational mechanisms effectively translate into measurable certification success. Through Partial Least Squares Structural Equation Modeling (PLS-SEM), the study assesses both direct and indirect effects among these constructs. By conceptualizing tryout performance as a mediating mechanism, the study clarifies how the implementation of digital assessment and competency management jointly contribute to policy outcomes. This integrated approach provides empirical evidence for evaluating policy effectiveness and offers strategic guidance for strengthening maritime certification systems under the CBA UKP framework.

A. Conceptualization and Measurement of Graduation Rates

Graduation rates are widely conceptualized as the proportion of students who complete an academic program within a defined time frame, typically measured as completion within a specified number of years after initial enrollment (Komasawa et al., 2022). In higher and professional education, this metric functions not only as an indicator of student achievement but also as a benchmark of institutional effectiveness and program quality. Particularly in professional certification systems, graduation or pass rates reflect whether educational processes successfully prepare candidates to meet competency standards required for workforce entry.

The measurement of graduation rates, however, is methodologically complex. Institutions may use cohort-tracking systems that follow a group of students over time, while others apply statistical modeling techniques to account for delayed completion, attrition, or nonlinear academic pathways (Komasawa et al., 2022; Kolagari et al., 2018). These methodological differences can significantly impact reported outcomes. For example, distinguishing between first-time students and transfer students is important because these groups may show different progression patterns (Arora & Al-Hattami, 2025). Additionally, attrition and repeated academic years influence completion statistics and must be carefully incorporated into analytical frameworks (Kolagari et al., 2018). From a public policy evaluation perspective, graduation rates can also be seen as outcome-based indicators of policy effectiveness. According to policy evaluation theory (Dunn; Anderson), effectiveness refers to the extent to which a policy achieves its intended goals. In the context of implementing computer-based assessments, graduation rates (Z) serve as the measurable outcome by which policy success is assessed. Therefore, graduation rates are not merely academic indicators but also tools for assessing whether assessment policies improve certification performance.

B. Theoretical Frameworks on Technology Utilization and Policy Efficiency

The Technology Acceptance Model (TAM) provides a foundational theoretical framework for understanding how users adopt and utilize technology in educational environments. TAM posits that perceived ease of use and perceived usefulness significantly shape users' attitudes toward a system, which subsequently influence behavioral intention and actual usage (Butler-Henderson & Crawford, 2020; Tian et al., 2018). In assessment contexts, perceived usefulness may relate to improved efficiency, clarity of instructions, immediate feedback, and enhanced control during examinations.

The Unified Theory of Acceptance and Use of Technology (UTAUT) extends TAM by incorporating constructs such as social influence and facilitating conditions, thereby acknowledging the institutional and environmental dimensions of technology adoption (Nardi & Ranieri, 2018; Kannan & Zapata-Rivera, 2022). Facilitating conditions such as infrastructure reliability, technical support, and adequate training are particularly critical in high-stakes professional assessment systems.

Within a policy implementation framework, technology utilization (X1) reflects the operational dimension of policy efficiency and responsiveness. Efficiency concerns whether digital systems optimize assessment processes and reduce procedural constraints. Responsiveness refers to a system's ability to accommodate user needs and adapt to context. Empirical studies grounded in TAM and UTAUT consistently report positive associations among technology acceptance, academic engagement, self-efficacy, and performance outcomes (Nardi & Ranieri, 2018; Yi et al., 2023). Therefore, the use of the CBA system can be interpreted as a proxy for the effectiveness of the policy instrument in practice.

C. Baseline Competency and Equity in Policy Implementation

Baseline competency is the minimum set of knowledge, skills, and abilities required to perform professional tasks effectively and safely. In competency-based education systems, these baseline standards serve as entry-level benchmarks that candidates must meet before moving on to certification stages. Implementing baseline competency usually involves standardized assessments aligned with professional competency frameworks (Klerk et al., 2018). From a policy evaluation standpoint, baseline competency (X2) closely aligns with equity and accountability. Equity means that all candidates have sufficient foundational capacity to compete fairly within the assessment system. Research shows that gaps in academic preparation and access to resources can affect certification performance (Costa & Chen, 2023; Webb et al., 2017). Systematically measuring and improving baseline competency helps institutions reduce structural inequalities that could otherwise skew policy outcomes. Additionally, accountability in policy implementation requires that certification results accurately reflect professional readiness. Ensuring alignment between competency benchmarks and assessment design improves validity and guarantees that graduation outcomes truly demonstrate minimum professional standards (Kaspar et al., 2016). Therefore, managing baseline competency is a key part of the human resource aspect of policy implementation.

D. Simulation-Based Tryouts and Policy Responsiveness

Simulation-based assessments and structured tryouts have become important preparatory strategies in professional education. Authentic scenarios help build familiarity and improve understanding within context, thereby increasing readiness for high-stakes exams (Kaspar et al., 2016). Formative feedback aids skill development and boosts self-efficacy (Huebner et al., 2023), while collaborative involvement promotes reflective evaluation and problem-solving skills (Scoular et al., 2020). Within a policy evaluation framework, Tryout Performance (Y) serves as a mediator, linking implementation inputs to outcome effectiveness. Tryouts improve responsiveness by allowing adaptive preparation and feedback cycles. They also enhance accountability by ensuring that final exam performance is preceded by structured rehearsals and competency checks. As an intermediary implementation tool, tryouts illustrate how technology use and baseline skills translate into measurable performance improvements.

E. Integrated Policy Evaluation Framework and Research Gap

The literature shows that graduation rates are complex indicators influenced by demographic factors and institutional approaches (Komasawa et al., 2022; Arora & Al-Hattami, 2025). Technology acceptance models explain the use of digital assessment (Butler, Henderson & Crawford, 2020; Kannan & Zapata-Rivera, 2022), while competency-based education focuses on profession-specific evaluation (Klerk et al., 2018). Simulation-based training further improves readiness and performance (Huebner et al., 2023; Scoular et al., 2020). However, these areas are often studied separately. Limited research exists that combines technology use (X1), baseline competency (X2), tryout performance (Y), and graduation rates (Z) within a single policy evaluation model.

II. Method

The implementation of Computer-Based Assessment (CBA) within professional certification systems represents not only a technological transition but also a strategic public policy intervention aimed at improving assessment integrity, standardization, and accountability for outcomes. In the maritime sector, the adoption of Computer-Based Assessment for the Seafarer Professional Examination (UKP Pelaut) is grounded in regulatory mandates issued by the Directorate General of Sea Transportation under the Ministry of Transportation, aligning with international competency frameworks such as the Standards of Training, Certification, and Watchkeeping (STCW). These regulations formalize CBA as an official assessment mechanism intended to enhance objectivity, transparency, and alignment with global maritime competency standards. Thus, the implementation of CBA UKP constitutes a policy instrument designed to strengthen certification quality and workforce readiness.

The primary objective of this policy is to ensure that professional certification processes are valid, reliable, and aligned with standardized competency benchmarks. Computer-based assessment systems (CBAS) have become increasingly central to professional education, particularly in disciplines where certification outcomes determine workforce eligibility and professional mobility. Empirical research consistently indicates that CBAS positively influences graduation rates and certification outcomes by enhancing performance accuracy, offering immediate feedback, and reducing examination-related anxiety (Wilson, 2018; Kaspar et al., 2016; Guimarães et al., 2017). Formative computer-based testing, for example, has been shown to significantly improve licensure examination performance in diagnostic medical sonography, thereby strengthening alignment between academic preparation and labor-market demands (Wilson, 2018). In nursing education, CBAS implementation has enabled systematic monitoring of competency attainment and learning outcomes, supporting reliable graduation and qualification standards (Kaspar et al., 2016; Guimarães et al., 2017). These findings support the policy rationale that digital assessment enhances institutional accountability and the credibility of certification.

In competency-driven sectors, assessment systems are closely linked to standardized professional frameworks that specify required skills and knowledge. Competency standards, such as STCW, demonstrate how structured benchmarks guide assessment design and certification processes to align with industry needs. Standardized competency frameworks shape the structure and content of testing, reinforcing consistency between educational outcomes and workplace expectations (Kaspar et al., 2016). Following these frameworks improves the validity, reliability, and professional relevance of assessments, thereby supporting successful certification (Grosseck et al., 2023; Ovtšarenko & Safiulina, 2025). As a result, integrating CBAS within competency-based systems is expected to enhance both exam integrity and professional readiness, fulfilling the normative goals of the CBA UKP policy.

However, despite its regulatory foundation and normative objectives, implementation challenges persist. The transition toward digital examination systems presents substantial constraints related to technological infrastructure, uneven digital literacy among faculty, and risks of academic dishonesty (Butler-Henderson & Crawford, 2020; Grosseck et al., 2023). Reports of increased cheating during unsupervised online examinations have raised serious concerns regarding assessment integrity in digital contexts (Castro et al., 2023). Technical disruptions during assessments, including interface limitations and navigation constraints, may negatively affect student scores and completion times (Cochran et al., 2020). Furthermore, institutions must invest in systematic training and support to ensure that both instructors and students can effectively adapt to digital assessment environments (Costa & Chen, 2023; Ceka & O'Geen, 2019). These implementation barriers indicate a potential gap between policy design and operational outcomes.

Low pass rates on professional certification exams highlight the need to assess whether the goals of CBA implementation are being met. Certification results are often affected by multiple factors, including examination anxiety (Kolagari et al., 2018), differences in familiarity with traditional versus digital assessments (Tian et al., 2018; Guimarães et al., 2017), and wider contextual elements such as socioeconomic status and access to technology (Costa & Chen, 2023; Webb et al., 2017). These factors suggest that adopting technology alone might not lead to higher graduation rates, raising important questions about the policy's effectiveness.

Familiarity with computer-based testing (CBT) platforms has been recognized as a crucial factor that influences exam performance and confidence. Consistent exposure to digital assessment tools is associated with lower anxiety, increased self-efficacy, and better academic outcomes (Nardi & Ranieri, 2018; Nagy & Korom, 2023; Kolagari et al., 2018). From a cognitive load perspective, familiarity decreases extraneous processing, allowing learners to concentrate more effectively on relevant cognitive tasks (Wang et al., 2022). Structured practice tests or mock exams serve as preparatory tools that foster this familiarity. Formative mock testing has demonstrated clear advantages, including reduced test anxiety and enhanced performance on licensure exams (Wilson, 2018). Additionally, structured mock assessments offer practical feedback, helping candidates improve their preparation strategies before final evaluations (Wang et al., 2022; Ponce et al., 2020). These mechanisms suggest that the success of CBA policy implementation depends not only on the availability of technology but also on skills management and organized rehearsal systems.

Although prior studies substantiate the individual contributions of CBAS, competency standards, digital familiarity, and mock examinations, these elements are frequently examined in isolation. Limited research evaluates the implementation of computer-based assessment as a policy system, with success measured by outcome indicators such as graduation rates. There is insufficient empirical modeling of how CBA utilization and baseline competency management interact through tryout mechanisms to influence certification outcomes within sector-specific training systems governed by regulatory mandates.

Therefore, this study positions itself as an outcome-based evaluation of the implementation of the computer-Based Assessment policy in the Seafarer Professional Examination (UKP Pelaut). By developing and empirically testing a mediated structural model linking CBA utilization, baseline competency and knowledge, tryout performance, and graduation rates, this research evaluates whether the policy's operational mechanisms effectively translate into measurable certification success. Through Partial Least Squares Structural Equation Modeling (PLS-SEM), the study assesses both direct and indirect effects among these constructs. By conceptualizing tryout performance as a mediating mechanism, the study clarifies how the implementation of digital assessment and competency management jointly contribute to policy outcomes. This integrated approach provides empirical evidence for evaluating policy effectiveness and offers strategic guidance for strengthening maritime certification systems under the CBA UKP framework.

A. *Conceptualization and Measurement of Graduation Rates*

Graduation rates are generally understood as the proportion of students who complete an academic program within a specified time, usually measured as completion within a certain number of years after initial enrollment (Komasawa et al., 2022). In higher and professional education, this metric serves not only as an indicator of student achievement but also as a benchmark for institutional effectiveness and program quality. Especially in professional certification systems, graduation or pass rates indicate whether educational processes effectively prepare candidates to meet workforce competency standards.

The measurement of graduation rates, however, is methodologically complex. Institutions may use cohort-tracking systems that follow a group of students over time, while others apply statistical modeling techniques to account for delayed completion, attrition, or nonlinear academic pathways (Komasawa et al., 2022; Kolagari et al., 2018). These methodological differences can significantly impact reported outcomes. For example, distinguishing between first-time students and transfer students is important because these groups may exhibit different progression patterns (Arora & Al-Hattami, 2025). Additionally, attrition and repeated academic years influence completion statistics and must be carefully incorporated into analytical frameworks (Kolagari et al., 2018). From a public policy evaluation perspective, graduation rates may also serve as outcome-based indicators of policy effectiveness. According to policy evaluation theory (Dunn; Anderson), effectiveness refers to the extent to which a policy achieves its intended objectives. In the context of computer-based assessment implementation, graduation rates (Z) represent the measurable outcome through which policy success can be assessed. Thus, graduation rates are not merely academic indicators but also tools for assessing whether assessment policies improve certification performance.

B. *Theoretical Frameworks on Technology Utilization and Policy Efficiency*

The Technology Acceptance Model (TAM) provides a foundational theoretical framework for understanding how users adopt and utilize technology in educational environments. TAM posits that perceived ease of use and perceived usefulness significantly shape users' attitudes toward a system, which subsequently influence behavioral intention and actual usage (Butler-Henderson & Crawford, 2020; Tian et al., 2018). In assessment contexts, perceived usefulness may relate to improved efficiency, clarity of instructions, immediate feedback, and enhanced control during examinations.

The Unified Theory of Acceptance and Use of Technology (UTAUT) extends TAM by incorporating constructs such as social influence and facilitating conditions, thereby acknowledging the institutional and environmental dimensions of technology adoption (Nardi & Ranieri, 2018; Kannan & Zapata-Rivera, 2022). Facilitating conditions such as infrastructure reliability, technical support, and adequate training are particularly critical in high-stakes professional assessment systems.

Within a policy implementation framework, technology utilization (X1) reflects the operational dimension of policy efficiency and responsiveness. Efficiency concerns whether digital systems optimize assessment processes and reduce procedural constraints. Responsiveness refers to a system's ability to accommodate user needs and to be contextually ready. Empirical studies grounded in TAM and UTAUT consistently report positive associations among technology acceptance, academic engagement, self-efficacy, and performance outcomes (Nardi & Ranieri, 2018; Yi et al., 2023). Therefore, the use of the CBA system can be interpreted as a proxy for how effectively the policy instrument functions in practice.

C. *Baseline Competency and Equity in Policy Implementation*

Baseline competency refers to the minimum set of knowledge, skills, and abilities required to perform professional tasks effectively and safely. In competency-based education systems, these baseline standards function as entry-level benchmarks that candidates must meet before advancing to certification stages. Operationalization of baseline competency typically involves standardized assessments aligned with professional competency frameworks (Klerk et al., 2018).

From a policy evaluation perspective, baseline competency (X2) closely relates to the equity and accountability dimension. Equity requires that all candidates have sufficient foundational capacity to compete fairly within the assessment system. Research indicates that disparities in academic preparation and access to resources may influence certification performance (Costa & Chen, 2023; Webb et al., 2017). By systematically measuring and strengthening baseline competency, institutions can mitigate structural inequalities that may otherwise distort policy outcomes.

Furthermore, accountability in policy implementation requires that certification outcomes accurately reflect professional readiness. Alignment between competency benchmarks and assessment design enhances validity and ensures that graduation outcomes genuinely represent minimum professional standards (Kaspar et al., 2016). Thus, baseline competency management constitutes a human resource dimension of policy implementation.

D. *Simulation-Based Tryouts and Policy Responsiveness*

Simulation-based assessments and structured tryouts have gained prominence as preparatory strategies in professional education. Authentic scenarios increase familiarity and contextual understanding, thereby enhancing readiness for high-stakes examinations (Kaspar et al., 2016). Formative feedback supports skill refinement and strengthens self-efficacy (Huebner et al., 2023), while collaborative engagement enhances reflective evaluation and problem-solving capacity (Scoular et al., 2020).

Within a policy evaluation framework, Tryout Performance (Y) serves as a mediating mechanism linking implementation inputs to outcome effectiveness. Tryouts enhance responsiveness by enabling adaptive preparation and feedback cycles. They also strengthen accountability by ensuring that final examination performance is preceded by structured rehearsal and competency verification.

By functioning as an intermediate implementation mechanism, tryouts operationalize how technology utilization and baseline competency translate into measurable performance improvements.

E. *Integrated Policy Evaluation Framework and Research Gap*

The literature demonstrates that graduation rates are multidimensional indicators influenced by demographic characteristics and institutional methodologies (Komasawa et al., 2022; Arora & Al-Hattami, 2025). Technology acceptance frameworks explain the utilization of digital assessment (Butler-Henderson & Crawford, 2020; Kannan & Zapata-Rivera, 2022), while competency-based education emphasizes profession-aligned evaluation (Klerk et al., 2018). Simulation-based preparation further enhances readiness and performance (Huebner et al., 2023; Scoular et al., 2020).

However, these streams are often examined independently. There remains limited research that integrates technology utilization (X1), baseline competency (X2), tryout performance (Y), and graduation rates (Z) within a unified policy implementation evaluation model.

III. Results and Discussion

This section presents the study's empirical findings in an objective manner. The results are organized according to the research framework, covering instrument testing, descriptive statistics, measurement model evaluation (outer model), structural model evaluation (inner model), and mediation analysis. Interpretation and theoretical discussion are intentionally reserved for the subsequent section.

A. *Instrument Testing (Pilot Study)*

Prior to full data analysis, a pilot test was conducted involving 30 respondents to evaluate the validity and reliability of the measurement instrument. The evaluation followed reflective measurement criteria in PLS-SEM using SmartPLS 4.0.

Convergent validity was assessed using outer loadings and Average Variance Extracted (AVE). The minimum acceptable outer loading threshold ranged between 0.60 and 0.70, and AVE values were required to exceed 0.50. Reliability was evaluated using Cronbach's Alpha (> 0.60) and Composite Reliability (> 0.60).

All indicators in the pilot phase met the required thresholds. No items were removed at this stage, and the instrument proceeded to full-sample analysis.

B. *Descriptive Statistics*

1) *Respondent Characteristics*

The total population consisted of 255 tryout participants in 2023, comprising 154 ANT I and 98 ATT I participants. Using the Taro Yamane formula with a 5% margin of error (d = 0.05), the required sample size was calculated as 155.73 and rounded to 156 respondents. The final dataset included 156 valid responses.

2) *Descriptive Results of Research Variables*

Four main constructs were analyzed descriptively:

- X1: Utilization of the CBA UKP Application
- X2: Baseline Competency & Knowledge
- Y: Try Out
- Z: Graduation Rate

Descriptive summaries for each construct are presented in Tables 1–4.

Table 1 summarizes respondents' perceptions of the CBA UKP application across nine indicators (X11–X19) using mean scores and standard deviations on a 1–5 scale. The mean values range from 3.62 to 4.29, indicating that responses generally cluster in the upper middle to high range. The highest mean appears on X19 (M = 4.29, SD = 0.63), while the lowest mean is X14 (M = 3.62, SD = 1.09). Standard deviations range from 0.63 to 1.23, indicating that some indicators have greater consensus than others. For example, X19 has relatively low dispersion (SD = 0.63), whereas X16 shows higher variability (SD = 1.23). Overall, the table provides item-level evidence on participants' ratings of the usefulness, effectiveness, and satisfaction with the CBA UKP system.

Table 1. Descriptive Statistics – Utilization of CBA UKP Application (X1)

Indicator	**Description**	**Mean**	**Std. Deviation**
X11	The CBA UKP application is easy to use for all users.	3.97	1.06
X12	The CBA UKP application is able to complete intended tasks properly.	3.76	1.19
X13	The CBA UKP application is effective in helping users achieve their objectives.	3.99	1.03
X14	The results produced by the CBA UKP application are accurate and reliable.	3.62	1.09
X15	The CBA UKP application functions properly and rarely experiences technical problems.	3.65	1.17
X16	The CBA UKP application is efficient in completing assigned tasks.	3.82	1.23
X17	I feel satisfied with the use of the CBA UKP application.	3.63	1.20
X18	Users perceive that the CBA UKP application is easy to use.	3.65	1.22
X19	Users perceive that the CBA UKP application effectively meets their needs.	4.29	0.63

Table 2 presents descriptive statistics for baseline competency and knowledge (X21–X29). Mean scores range from 3.72 to 4.53, suggesting that participants tend to report moderately high levels across most competency-related statements. The highest mean is recorded for X22 (M = 4.53, SD = 0.66), while the lowest mean is X23 (M = 3.72, SD = 0.71). The standard deviations range from 0.62 to 0.87, indicating generally moderate response variability. Several indicators show relatively strong central tendency with lower dispersion, such as X24 (M = 4.37, SD = 0.62). By reporting item-level means and dispersion, this table clarifies which aspects of baseline capability (knowledge, technical skills, interpersonal skills, commitment, and motivation) are rated more consistently, and which show wider differences among respondents.

Table 2. Descriptive Statistics – Baseline Competency & Knowledge (X2)

Indicator	**Description**	**Mean**	**Std. Deviation**
X21	I have sufficient knowledge about the tasks or work I must perform.	4.33	0.73
X22	I understand theories and concepts relevant to my tasks or work.	4.53	0.66
X23	I have good knowledge of rules and procedures related to my tasks or work.	3.72	0.71
X24	I possess the technical skills required to complete my tasks or work.	4.37	0.62
X25	I have good interpersonal skills to collaborate with others.	4.26	0.73
X26	I am able to solve problems arising in my tasks or work effectively.	3.97	0.78

Indicator	Description	Mean	Std. Deviation
X27	I have a positive attitude toward my tasks or work.	3.78	0.83
X28	I am committed to completing my tasks or work properly.	3.84	0.87
X29	I am motivated to learn and develop in my tasks or work.	4.36	0.73

Table 3 reports descriptive statistics for the try-out construct (Y1–Y9), covering perceived ability, readiness, confidence, emotional state, and strategy during the try-out. Mean scores range from 3.72 to 4.41, with most items above 4.00 except Y1 and Y4. The highest mean is Y3 (M = 4.41, SD = 0.70), followed closely by Y6 (M = 4.39, SD = 0.59) and Y2 (M = 4.37, SD = 0.67). The lowest mean is Y1 (M = 3.72, SD = 0.91). Standard deviations range from 0.59 to 0.98, suggesting moderate variability across items, with Y9 showing the widest dispersion (SD = 0.98). Notably, Y5 ("I feel anxious when facing the try out") is an affective item that may require attention in coding (e.g., reverse-coding) depending on the measurement model specification.

Table 3. Descriptive Statistics – Try Out (Y)

Indicator	Description	Mean	Std. Deviation
Y1	I can complete try out questions correctly.	3.72	0.91
Y2	I understand the material tested in the try out.	4.37	0.67
Y3	I am able to answer try out questions within the allotted time.	4.41	0.70
Y4	I feel confident when facing the try out.	3.99	0.66
Y5	I feel anxious when facing the try out.	4.33	0.69
Y6	I feel mentally and physically prepared when facing the try out.	4.39	0.59
Y7	I use effective learning strategies in preparing for the try out.	4.22	0.70
Y8	I use appropriate answering strategies during the try out.	4.17	0.90
Y9	I use effective time management strategies during the try out.	4.23	0.98

Table 4 presents descriptive statistics for perceptions of graduation rates (Z1–Z9), reflecting respondents' views on certification outcomes, institutional performance, learning influence, and contextual factors. Mean scores range from 3.56 to 4.46, indicating that responses vary from moderate to high levels across indicators. The highest mean is Z9 (M = 4.46, SD = 0.90), while the lowest mean is Z1 (M = 3.56, SD = 0.89). Several indicators show high mean values above 4.20, such as Z2 (M = 4.22, SD = 0.88) and Z8 (M = 4.34, SD = 0.82). Standard deviations range from 0.76 to 0.91, indicating moderate dispersion and relatively consistent perceptions across respondents. This table functions as an item-level summary of how participants evaluate graduation/pass-rate conditions and perceived drivers of certification success.

Table 4. Descriptive Statistics – Graduation Rate (Z)

Indicator	Description	Mean	Std. Deviation
Z1	I believe the percentage of participants who pass the Seafarer Expertise Examination Level I (ANT I & ATT I) is quite high when they participate in the try out.	3.56	0.89
Z2	I believe the percentage of participants achieving the minimum required score in the Seafarer Expertise Examination Level I (ANT I & ATT I) is adequate.	4.22	0.88
Z3	I believe the percentage of participants accepted in their chosen maritime field is relatively high.	4.21	0.91
Z4	I believe the graduation rate at STIP is better compared to other educational institutions.	4.13	0.80
Z5	I believe the graduation rate of participants in recent years has increased after participating in the try out.	4.06	0.76
Z6	I believe graduation rates across departments or study programs at our institution are balanced.	4.12	0.85
Z7	The learning methods used at our institution are very effective in improving graduation rates.	3.88	0.81
Z8	Learning at our institution influences graduation rates.	4.34	0.82
Z9	I believe that participants' socioeconomic conditions do not significantly affect graduation rates at STIP.	4.46	0.90

C. *Measurement Model Evaluation (Outer Model)*

1) *Convergent Validity*

Convergent validity was assessed using outer loadings and AVEs. All reported outer loadings exceeded 0.70, satisfying the recommended minimum criteria.

Table 5 reports the validity results for the measurement model at two levels: (1) first order (indicator level) and (2) second-order (dimension level). At the indicator level, each item (e.g., X11–X19; X21–X29; Y1–Y9; Z1–Z9) is evaluated based on its outer loading, which reflects how strongly it represents its latent construct. At the dimension level, each sub-dimension (e.g., Usefulness, Effectiveness, User Satisfaction) is assessed using a dimension loading, which indicates how well the dimension contributes to the higher-order construct. In reflective PLS-SEM, loadings ≥ 0.70 are commonly considered strong, while values in the 0.60–0.70 range may still be acceptable depending on context. Because all indicators and dimensions are labeled "Valid," the table shows that each construct meets the required validity criteria and can be retained for structural model testing. All loadings were above 0.70, indicating satisfactory convergent validity.

Table 5. First Order and Second-Order Validity Testing – Utilization of CBA UKP Application (X1)

Construct	Dimension	Indicator	Dimension Loading	Indicator Loading	Status
Utilization of CBA UKP (X1)	Usefulness	X11	0.898	0.808	Valid
		X12		0.790	Valid
		X13		0.952	Valid
	Effectiveness	X14	0.931	0.860	Valid
		X15		0.896	Valid
		X16		0.885	Valid
	User Satisfaction	X17	0.957	0.841	Valid
		X18		0.901	Valid
		X19		0.857	Valid
Baseline Competency & Knowledge (X2)	Knowledge	X21	0.893	0.952	Valid
		X22		0.867	Valid
		X23		0.836	Valid
	Skills	X24	0.913	0.798	Valid
		X25		0.911	Valid
		X26		0.898	Valid
	Attitude	X27	0.849	0.923	Valid
		X28		0.839	Valid
		X29		0.905	Valid
Try Out (Y)	Ability	Y1	0.935	0.761	Valid
		Y2		0.900	Valid
		Y3		0.887	Valid
	Readiness	Y4	0.949	0.761	Valid
		Y5		0.838	Valid
		Y6		0.884	Valid
	Strategy	Y7	0.890	0.731	Valid
		Y8		0.923	Valid
		Y9		0.908	Valid
Graduation Rate (Z)	Percentage	Z1	0.941	0.901	Valid
		Z2		0.903	Valid
		Z3		0.904	Valid
	Comparison	Z4	0.932	0.903	Valid
		Z5		0.944	Valid
		Z6		0.941	Valid
	Factor	Z7	0.950	0.870	Valid
		Z8		0.923	Valid
		Z9		0.925	Valid

2) Discriminant Validity

Discriminant validity was evaluated using cross-loadings and the Fornell–Larcker criterion. Each construct's square root of AVE exceeded its correlations with other constructs. Cross-loading values for each indicator were higher for its associated construct than for the other constructs.

3) Reliability

Reliability testing indicated that all constructs achieved Composite Reliability values above 0.60 and Cronbach's Alpha values above 0.60, meeting recommended thresholds.

Table 6 presents the internal consistency reliability results for all latent constructs using Cronbach's Alpha and Composite Reliability. Reliability testing evaluates whether the indicators within each construct consistently measure the same concept. In PLS-SEM, values above 0.70 are generally considered acceptable for confirmatory research. The results show that all constructs exceed this threshold, confirming strong internal consistency. Specifically, Utilization of CBA UKP Application (X1) demonstrates Cronbach's Alpha = 0.937 and Composite Reliability = 0.947. Baseline Competency & Knowledge (X2) achieves Cronbach's Alpha = 0.919 and Composite Reliability = 0.933, while Try Out (Y) also records

Vol. 9, Issue 1, June 2026, pp. 01-15 e-ISSN: 2654-5667

Cronbach's Alpha = 0.919 and Composite Reliability = 0.933. Graduation Rate (Z) shows the highest reliability values (Cronbach's Alpha = 0.957; Composite Reliability = 0.963). Overall, these results confirm that the measurement items are reliable and suitable for subsequent structural model testing.

Table 6. Reliability Results

Construct	Cronbach's Alpha	Composite Reliability	Rule of Thumb	Description
Utilization of CBA UKP Application (X1)	0.937	0.947	> 0.70	Reliable
Baseline Competency & Knowledge (X2)	0.919	0.933	> 0.70	Reliable
Try Out (Y)	0.919	0.933	> 0.70	Reliable
Graduation Rate (Z)	0.957	0.963	> 0.70	Reliable

D. *Structural Model Evaluation (Inner Model)*

1) *Direct Effects*

Path coefficients (β), t-statistics, and p-values were examined to evaluate direct relationships.

Table 7. Direct Effects (Path Coefficients)

Hypothesis	Effect	Original Sample (O)	T Statistics (|O/STDEV|)	P Values	Description
H1	Utilization of CBA UKP Application (X1) → Try Out (Y)	0.686	11.163	0.000	Positive and significant
H2	Baseline Competency & Knowledge (X2) → Try Out (Y)	0.388	6.108	0.000	Positive and significant
H3	Utilization of CBA UKP Application (X1) → Graduation Rate (Z)	0.324	1.669	0.048	Positive and significant
H4	Baseline Competency & Knowledge (X2) → Graduation Rate (Z)	0.310	3.611	0.000	Positive and significant
H5	Try Out (Y) → Graduation Rate (Z)	0.364	1.682	0.047	Positive and significant

The path coefficient from X1 to Y was 0.686 with T = 11.163 and p = 0.000. The path from X2 to Y was 0.388 with T = 6.108 and p = 0.000. The path from X1 to Z was 0.324 (positive direction). The path from X2 to Z yielded T = 3.611, p = 0.000.

2) *Coefficient of Determination (R^2)*

R-square values were examined to determine the proportion of variance explained by the model.

Table 8 presents the R-square (R^2) values for the endogenous variables in the structural model typically Try Out (Y) and Graduation Rate (Z). R^2 indicates the proportion of variance explained by the predictor variables in the model. For example, the R^2 for Try Out (Y) reflects how much of tryout performance is explained by CBA utilization (X1) and baseline competency (X2). Meanwhile, the R^2 for Graduation Rate (Z) reflects how much of the variation in graduation outcomes is explained by X1, X2, and Y (the mediator). In behavioral and education research, R^2 values are often interpreted using benchmarks such as 0.75 (strong), 0.50 (moderate), and 0.25 (weak). Therefore, Table 8 is essential for demonstrating whether the model has meaningful explanatory power and whether the proposed predictors collectively provide a substantial explanation of certification outcomes. R^2 interpretation followed established benchmarks: 0.75 (strong), 0.50 (moderate), 0.25 (weak).

Table 8. R-Square Values

Endogenous Variable	R Square	Q Square
Try Out (Y)	0.865	0.519
Graduation Rate (Z)	0.787	0.579

3) *Structural Model Figures*

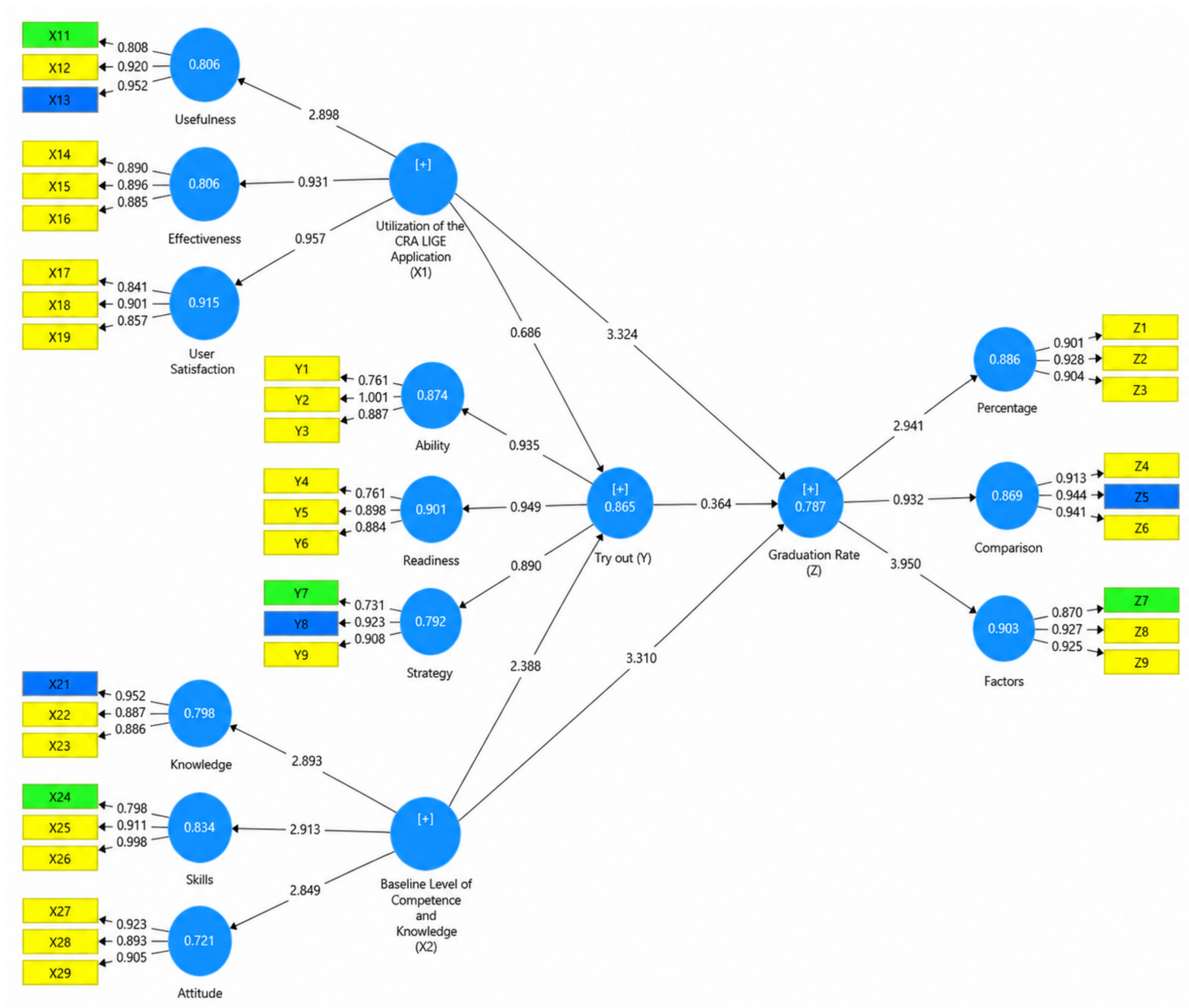

Fig. 1. Outer Model (SmartPLS Output, 2024) Caption: Reflective measurement model showing relationships between latent constructs and indicators.

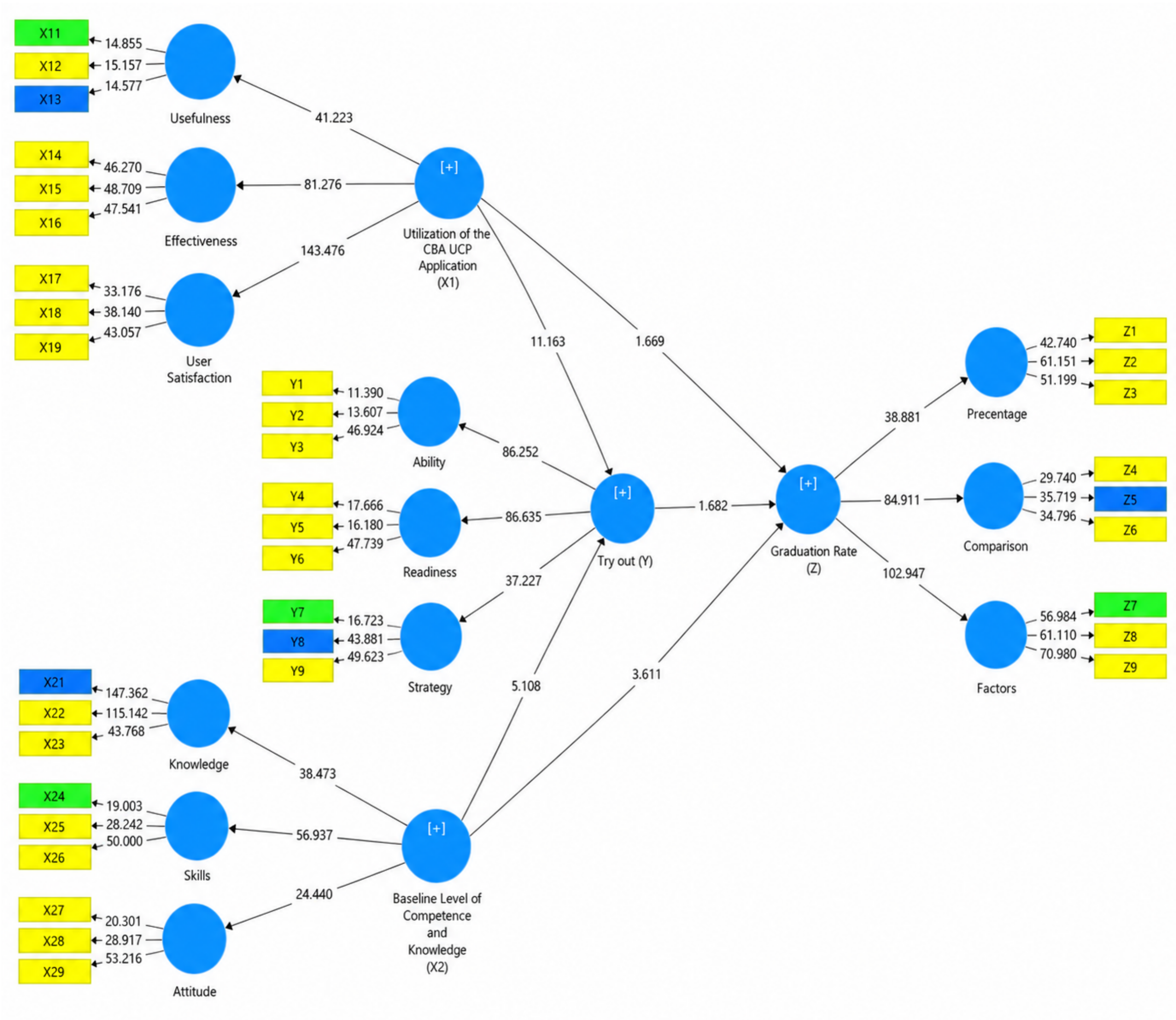

Fig. 2. Inner Model (Structural Model)

Caption: Structural relationships among CBA Utilization (X1), Baseline Competency (X2), Try Out (Y), and Graduation Rate (Z), including direct and indirect paths.

E. Mediation Analysis

Mediation analysis was conducted using bootstrapping procedures to examine indirect effects.

Table 9. Indirect Effects (Mediation Results)

Hypothesis	Effect	Original Sample (O)	T Statistics (|O/STDEV|)	P Values	Remarks
H6	Utilization of CBA UKP Application (X1) → Try Out (Y) → Graduation Rate (Z)	0.250	1.693	0.046	Valid
H7	Baseline Competency & Knowledge (X2) → Try Out (Y) → Graduation Rate (Z)	0.141	1.732	0.042	Valid

Try Out (Y) served as a mediating variable between CBA Utilization (X1) and Graduation Rate (Z), as well as between Baseline Competency (X2) and Graduation Rate (Z). Bootstrapping confirmed the statistical significance of the indirect effects. Overall, the empirical findings show statistically significant direct and mediated relationships among the studied variables within the proposed structural model. No interpretive conclusions are drawn in this section, as analytical implications are discussed in the Discussion section.

F. Evaluation of CBA Policy Implementation Effectiveness

This study was designed as an outcome-based evaluation of the implementation of the Computer-Based Assessment (CBA) policy in the Seafarer Professional Examination (UKP Pelaut). The findings provide empirical evidence regarding the effectiveness and operational performance of this policy instrument.

The strong and statistically significant effect of CBA Utilization (X1) on Try Out performance ($\beta = 0.686$; $p < 0.05$) indicates that the digital assessment system functions effectively at the operational level. From a policy implementation perspective, this result reflects the efficiency and responsiveness dimensions of policy performance. When participants perceive the system as useful, reliable, and accessible, their readiness for high-stakes examinations improves substantially. This suggests that, at least in terms of technological functionality and user engagement, the CBA UKP policy has been implemented with measurable success.

However, policy effectiveness cannot be measured solely through operational usage. The direct impact of CBA utilization on Graduation Rate ($\beta = 0.324$; $p < 0.05$) shows that digital assessment has a positive effect on certification outcomes, but the effect size remains moderate. This suggests that, although the policy tool is functioning, it alone is not enough to ensure optimal graduation outcomes. Therefore, the policy can be considered partially effective rather than fully transformative. Baseline Competency & Knowledge (X2) also has significant effects on both Try Out performance ($\beta = 0.388$; $p < 0.05$) and Graduation Rate ($T = 3.611$; $p < 0.05$). This highlights the importance of equity and accountability in policy. The CBA policy presumes that candidates possess sufficient foundational skills. When baseline competency is strong, policy results improve. When competency gaps exist, graduation outcomes may stay below expectations. This indicates that the success of implementation depends on parallel human resource competency management.

The mediation analysis further clarifies how implementation works. The Try Out performance functions as an intermediate link connecting policy inputs (digital systems and competency management) to outcome effectiveness (graduation rates). The presence of significant indirect effects confirms that policy success is supported by structured preparatory mechanisms. This shows that the effectiveness of the CBA UKP policy increases when simulation-based rehearsal systems are institutionalized. Overall, the findings indicate that the CBA UKP policy has achieved moderate to high effectiveness, especially in operational efficiency and responsiveness. However, achieving full effectiveness in improving outcomes requires additional institutional support mechanisms.

G. Alignment with Policy Objectives and Literature

The positive relationship between digital assessment utilization and examination performance aligns with prior research emphasizing the benefits of standardized digital testing in professional education (Kaspar et al., 2016; Wilson, 2018). These studies support the rationale underlying the CBA policy: digital systems enhance objectivity, monitoring, and alignment with competency standards.

However, cross-sector comparisons indicate that technological policy instruments produce varying levels of impact depending on institutional readiness (Thi & Hoang, 2024). The current findings suggest that in contexts where infrastructure and structured tryouts are systematically implemented, the CBA policy yields measurable performance gains. This reinforces the view that policy effectiveness is conditional upon implementation capacity.

The significant contribution of baseline competency is consistent with competency-based education frameworks (Grosseck et al., 2023; Schrempft et al., 2021). From a policy standpoint, this finding underscores the need for certification reforms to integrate both technological and human resource strategies. A digital assessment policy that does not simultaneously strengthen competency management risks underperforming relative to its intended objectives.

The mediating role of Try Out performance further confirms the importance of formative mechanisms, as recommended in digital assessment research (Tian et al., 2018; Costa & Chen, 2023). In policy terms, tryouts operationalize responsiveness and accountability by ensuring that participants receive structured preparation prior to final evaluation.

H. Assessment of Policy Effectiveness

Based on the empirical findings, the implementation of the CBA UKP policy can be evaluated as follows:

- Effectiveness: The policy positively influences graduation outcomes, but the impact is partially mediated and not solely determined by digital utilization. Therefore, effectiveness is significant but not absolute.
- Efficiency: High influence of CBA utilization on tryout readiness suggests strong operational efficiency of the digital system.
- Responsiveness: Structured tryouts enhance adaptive preparation and strengthen policy responsiveness.
- Equity: Variations in baseline competency suggest that unequal foundational preparation may influence outcomes, underscoring the need to ensure equitable readiness.
- Accountability: Integrating competency standards and measurable graduation outcomes strengthens certification accountability.

Thus, the CBA UKP policy can be categorized as partially effective and operationally functional, but requiring strategic refinement to optimize outcome consistency.

I. Policy Improvement Implications

The findings generate several evidence-based policy recommendations:

- Institutionalize mandatory structured tryouts. Tryout mechanisms should be formally integrated into the certification process as mandatory preparatory stages, due to their role in improving policy outcomes.
- Strengthen Competency Management Systems Pre-examination competency reinforcement programs should be standardized to ensure equal preparedness among participants. This includes diagnostic baseline testing and targeted training remediation.
- Enhance continuous digital system optimization. While operationally effective, ongoing monitoring of the CBA system's usability, infrastructure stability, and technical support is essential to maintain efficiency and performance responsiveness.
- Implement periodic policy outcome monitoring. Graduation rate data should be systematically analyzed across cohorts to identify implementation gaps and inform iterative policy adjustments.
- Integrate Human Resource and Digital Policy Components
  The CBA policy should not function as a stand-alone technological reform. Instead, it must be integrated with broader maritime competency management strategies aligned with STCW standards.

J. Limitations and Future Policy Evaluation Directions

The cross-sectional design limits the ability to assess policy impact over time (Hamhuis et al., 2020; Park, 2024). Future research should use longitudinal policy evaluation methods to track certification trends across multiple cohorts. Additionally, relying on self-reported perceptions may introduce response bias. Using mixed-method approaches that include qualitative analysis of policy implementation (Castro et al., 2023) would offer deeper insights into institutional dynamics and participants' experiences. Expanding the evaluation model to include variables such as institutional governance quality, instructor competence, and resource allocation would enable a more comprehensive assessment of policy implementation effectiveness.

IV. Conclusion

This study evaluates the outcome-based implementation of the Computer-Based Assessment (CBA) policy in the Seafarer Professional Certification Examination, examining the links between CBA usage, baseline competence, tryout performance, and graduation rates. Results show that while the CBA system enhances tryout performance and directly supports graduation outcomes, and baseline competence strongly influences readiness and success, the policy is only partially effective since technology adoption alone cannot ensure optimal results. Tryout performance mediates these relationships, underscoring accountability and responsiveness, but overall effectiveness depends on integrating digital tools with competency management and structured preparation. Policy recommendations include strengthening training and technical support, establishing systematic competency programs, mandating structured tryouts, and conducting regular monitoring to align certification outcomes with workforce standards.

References

Ambiyar, A., Panyahuti, P., Devega, A. T., & Islami, S. (2024). Web and Android-based test application development and its implementation on final semester examination. *Joiv International Journal on Informatics Visualization, 8*(2), 768. https://doi.org/-10.62527/joiv.8.2.2120

Arora, B., & Al-Hattami, A. (2025). Technological tool for formative assessment in higher education: ZipGrade. *Canadian Journal of Learning and Technology, 50*(2), 1–18. https://doi.org/10.21432/cjlt28759

Butler-Henderson, K., & Crawford, J. (2020). A systematic review of online examinations: A pedagogical innovation for scalable authentication and integrity. *Computers & Education, 159*, 104024. https://doi.-org/10.1016/j.compedu.2020.104024

Castro, M., Manoharan, S., Speidel, U., Ye, X., & Zu, J. (2023). Observations of cheating behaviors in online examinations and tools for mitigation (pp. 1–7). https://doi.org/10.1109/fie58773.2023.10343363

Ceka, B., & O'Geen, A. J. (2019). Evaluating student performance on computer-based versus handwritten exams: Evidence from a field experiment in the classroom. *PS: Political Science & Politics, 52*(4), 757–762. https://doi.org/10.1017/S10490965190009-1X

Cochran, G. L., Foster, J., Klepser, D. G., Dobesh, P. P., & Dering-Anderson, A. M. (2020). The impact of eliminating backward navigation on computerized examination scores and completion time. *American Journal of Pharmaceutical Education, 84*(12), ajpe8034. https://doi.org/10.5688/ajpe8034

Costa, D. R., & Chen, C.-W. (2023). Exploring the relationship between process data and contextual variables among Scandinavian students on PISA 2012 mathematics tasks. *Large-Scale Assessments in Education, 11*(1). https://doi.org/10.1186/s40536-023-00155-x

Cutumisu, M. (2018). The informational value of feedback choices for performance and revision in a digital assessment game. *Interactive Technology and Smart Education, 15*(4), 363–380. https://doi.org/10.11-08/ITSE-03-2018-0017

Grosseck, G., Bran, R., & Țîru, L. G. (2023). Digital assessment: A survey of Romanian higher education teachers' practices and needs. *Education Sciences, 14*(1), 32. https://doi.org/10.3390/educsci14010032

Guimarães, B., Ribeiro, J. C., Cruz, B., et al. (2017). Performance equivalency between computer-based and traditional pen-and-paper assessment: A case study in clinical anatomy. *Anatomical Sciences Education, 11*(2), 124–136. https://doi.org/10.10-02/ase.1720

Hamhuis, E. R., Glas, C. A. W., & Meelissen, M. R. (2020). Tablet assessment in primary education: Are there performance differences between TIMSS paper-and-pencil and tablet tests among Dutch grade-four students? *British Journal of Educational Technology, 51*(6), 2340–2358. https://doi.org/10.1111/bjet.12914

Huebner, B., Jackson, B., Danzl, M. M., & Pitt, J. (2023). Exam review versus categorical feedback: Which improves scores by PDoctor of physical Therapy students? *Journal of Physical Therapy Education, 38*(1), 37–43. https://doi.org/10.1097/JTE.00000000-00000322

Kannan, P., & Zapata-Rivera, D. (2022). Facilitating the use of data from multiple sources for formative learning in the context of digital assessments. *Frontiers in Education, 7*. https://doi.org/10.3389/feduc.2022.91-3594

Kaspar, R., Döring, O., Wittmann, E., et al. (2016). Competencies in geriatric nursing: Empirical evidence from a computer-based large-scale assessment calibration study. *Vocations and Learning, 9*(2), 185–206. https://doi.org/10.1007/s12186-015-9147-y

Klerk, S. D., Veldkamp, B. P., & Eggen, T. J. H. M. (2018). The design, development, and validation of a multimedia-based performance assessment for credentialing confined space guards. *Behavior-metrika, 45*(2), 565–596. https://doi.org/10.1007/s4-1237-018-0064-x

Kolagari, S., Modanloo, M., Rahmati, R., Sabzi, Z., & Ataee, A. J. (2018). The effect of computer-based tests on nursing students' test anxiety: A quasi-experimental study. *Acta Informatica Medica, 26*(2), 115–118. https://doi.org/10.5455/aim.2018.26.115-118

Komasawa, N., Terasaki, F., Takitani, K., Lee, S.-W., Kawata, R., & Nakano, T. (2022). Comparison of younger and older medical student performance outcomes: A retrospective analysis in Japan. *Medicine, 101*(45), e31392. https://doi.org/10.1097/-MD.0000000000031392

Molina, L. F. M., Durán, W. F., & Valencia, C. A. (2024). What matters the most: The assessment method or individual skills in marketing research learning? *Journal of International Education in Business, 17*(2), 283–303. https://doi.org/10.1108/JIEB-07-2023-0043

Nagy, M. T., & Korom, E. (2023). Measuring scientific reasoning of fourth graders: Validation of the Science-K inventory in paper-based and computer-based testing environments. *Journal of Baltic Science Education, 22*(6), 1050–1062. https://doi.org/10.-33225/jbse/23.22.1050

Nardi, A., & Ranieri, M. (2018). Comparing paper-based and electronic multiple-choice examinations with personal devices: Impact on students' performance, self-efficacy, and satisfaction. *British Journal of Educational Technology, 50*(3), 1495–1506. https://-doi.org/10.1111/bjet.12644

Ovtšarenko, O., & Safiulina, E. (2025). Computer-driven assessment of weighted attributes for e-learning optimization. *Computers, 14*(4), 116. https://doi.-org/10.3390/computers14040116

Park, E. (2024). The exploration of EFL preservice teachers' self-perceived importance of assessment literacy. *Language Teaching Research Quarterly, 45*, 22–38. https://doi.org/10.32038/ltrq.2024.45.02

Ponce, H. R., Mayer, R. E., & Loyola, M. S. (2020). Effects on test performance and efficiency of technology-enhanced items. *Journal of Educational Computing Research, 59*(4), 713–739. https://doi.org/10.-1177/0735633120969666

Schrempft, S., Piumatti, G., Gerbase, M. W., & Baroffio, A. (2021). Pathways to performance in undergraduate medical students. *Advances in Health Sciences Education, 26*(5), 1537–1554. https://doi.org/10.10-07/s10459-021-10059-6

Scoular, C., Eleftheriadou, S., Ramalingam, D., & Cloney, D. (2020). Comparative analysis of student performance in collaborative problem solving. *Australian Journal of Education, 64*(3), 282–303. https://doi.org/10.-1177/0004944120957390

Thi, B. V., & Hoang, H. (2024). Digital competency: Perspectives of teacher students and primary teachers. *Multidisciplinary Reviews, 7*(12), 2024216. https://-doi.org/10.31893/multirev. 2024216

Tian, T., DeMara, R. F., & Gao, S. (2018). Efficacy and perceptions of assessment digitization. *Computer Applications in Engineering Education, 27*(2), 419–429. https://doi.org/10.1002/cae.22086

Wang, J.-F., Wang, T.-H., & Huang, C.-H. (2022). Investigating students' answering behaviors in a computer-based mathematics algebra test. *Behavioral Sciences, 12*(8), 293. https://doi.org/10.3390/bs1208-0293

Webb, L., Clough, J., O'Reilly, D., Wilmott, D., & Witham, G. (2017). The utility and impact of ICT for pre-registration nurse education. *Nurse Education Today, 48*, 160–171. https://doi.org/10.1016/j.nedt.2016.10.-007

Wilson, M. (2018). The impact of formative computer-based testing on learners' anxiety and performance. *Journal of Diagnostic Medical Sonography, 34*(6), 425–433. https://doi.org/10.1177/8756479318799338

Yi, S., Zhu, M., & Sparks, J. R. (2023). Exploring the role of process data analysis in understanding student performance. *Journal of Educational Computing Research, 61*(5), 1096–1120. https://doi.org/10.1177/-07356331221138734