International Journal of Education, Vocational and Social Science
https://e-journal.citakonsultindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

The Effect Of The Flipped Classroom Model On Critical Thinking Skills Reviewed From Self-Efficacy By Controlling Students' Initial Literacy And Numeracy Skills

I Dewa Putu Mas Dwipayana[1]
[1]Universitas Pendidikan Ganesha, Singaraja, Indonesia
I Nyoman Jampel[2]
[2]Universitas Pendidikan Ganesha, Singaraja, Indonesia
I Gede Ratnaya[3]
[3]Universitas Pendidikan Ganesha, Singaraja, Indonesia

Correspondence Author: *mas.dwipayana@student.undiksha.ac.id*

Abstract

This study aims to analyze in depth the influence of the Flipped Classroom learning model (consisting of Standard Flipped Classroom and Flipped Mastery Classroom) compared to conventional models on students' critical thinking skills in Informatics subjects, reviewed from the level of self-efficacy (high and low) by controlling students' initial literacy and numeracy skills as covariates. This research was carried out at SMAN 1 Ubud with a factorial design of 2 × 3 using a two-path covariance analysis (ANCOVA). The research population includes all students of class X for the 2025/2026 school year as many as 324 students. The sample was selected using a simple random sampling technique based on the class (intake group) which produced 3 treatment classes (108 students), then grouped based on the highest and lowest 27% of self-efficacy scores so that 60 students were obtained as an extreme subject analysis sample. Critical thinking ability data were measured by an integrated reasoning multiple-choice test (38 valid items, KR-20 = 0.954), self-efficacy was measured by a Likert scale questionnaire (30 valid items, Alpha Cronbach = 0.994), and early literacy-numeracy ability was measured by a diagnostic test (33 valid items, KR-20 = 0.982). The results showed that: (1) There was a significant difference in students' critical thinking skills between learning models after controlling for literacy and numeracy early ($F = 12826.750$; $p < 0.001$), where Flipped Mastery Classroom produced the highest achievement; (2) There was a significant effect of the level of self-efficacy on critical thinking skills ($F = 7901.998$;

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

p < 0.001); (3) There was a significant interaction effect between the learning model and self-efficacy on critical thinking skills (F = 416.106; p < 0.001); and (4) Controlling literacy and early numeracy is statistically proven to increase the precision and internal validity of the treatment estimation model. These findings confirm that the integration of adaptive Flipped Mastery strategies is very effective in facilitating independence, competency completion, and high-level thinking (HOTS) in the era of the Independent Curriculum.

Keywrod: Critical Thinking, Flipped Classroom, Flipped Mastery, Informatics, Literacy and Numeracy, Self-Efficacy

INTRODUCTION

The demands of the world of education in the 21st century have undergone a very fundamental paradigm shift along with the rapid flow of digitalization, artificial intelligence, and globalization of the information ecosystem. Education is no longer placed simply as a vehicle for knowledge transmission centered on teachers, but as a cognitive reconstruction process that equips students with higher-order thinking skills or Higher Order Thinking Skills (HOTS). The 4C skills, namely critical thinking, creativity, collaboration, and communication, have been internationally agreed as the main foundation of human resource capabilities to be able to adapt and solve complex unstructured problems in the modern era (Erlin et al., 2021). In this context, critical thinking skills are the central axis that determines the extent to which individuals are able to analyze, evaluate, filter information, and make logical and accountable inferences in the midst of a tsunami of digital data (Facione, 2020).

However, the pedagogical reality in the field shows a fairly sharp disparity between curricular ideals and the real competency achievements of students in Indonesia. Periodic reports from the Programme for International Student Assessment (PISA) organized by the OECD consistently place Indonesian students' reasoning skills in the bottom position globally. The results of PISA 2018 recorded the average score of Indonesian students in the field of reading at 371, mathematics 379, and science at 389, which placed Indonesia in 64th place out of 65 countries (Sälzer & Roczen, 2018). Furthermore, the 2022 PISA results show that Indonesia's science position is ranked 67 out of 81 countries with an average score of 383 (OECD, 2023). The majority of students in Indonesia are only able to complete cognitive tasks at Level 1 and Level 2 which are direct recall and routine procedures, and experience great difficulties when faced with

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

high-level problems that require data interpretation, contextual analysis, and rational argument evaluation.

As a transformative step nationally, the Government of the Republic of Indonesia launched the Independent Curriculum and National Assessment which places the development of Critical Reasoning character and strengthening literacy and numeracy as top priorities. The subject of Informatics is present in the Independent Curriculum not only as a mastery of information technology instruments, but as a discipline that integrates Computational Thinking, algorithmic logic, data analysis, and the social impact of technology. Informatics learning requires students to be familiar with systematic problem solving, problem decomposition, abstraction, and pattern recognition. Therefore, mastery of Informatics requires a very intensive level of critical reasoning.

Although the Independent Curriculum framework has been designed progressively, the implementation of learning at the education unit level still often clashes with conventional patterns. At SMAN 1 Ubud, based on data from the 2024 Education Report Report, there was a decrease in achievement scores in the competencies of "Knowing" and "Critical Reasoning" compared to 2023. The results of interviews and initial observations indicate that the Informatics learning process at SMAN 1 Ubud is still trapped in the dominance of lecture methods, rigid procedural instructions, and the lack of space for independent exploration based on open questions. Students tend to be passive consumers of information that depend on the teacher's explanations in class, so when faced with algorithmic case studies or complex data analysis, students encounter serious obstacles in building structured arguments and evaluating alternative solutions. Based on these considerations, this study aims to analyze the influence of the flipped classroom model on students' critical thinking skills in informatics subjects by considering self-efficacy as a variable that affects the effectiveness of the strategy. In particular, this study was designed to find out whether the use of the flipped classroom learning model can improve students' critical thinking skills, which include the ability to analyze, evaluate, infer logically, and apply learning concepts in real life.

RESEARCH METHODS

This study uses a quantitative approach with a quasiexperimental design with a Factorial 2 × 3 pattern. The use of pseudo-experiments was chosen because the research subjects were bound to natural classes that had been formed by the school (intake groups), so it was not

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

possible to carry out pure individual randomization without interfering with the academic schedule of SMAN 1 Ubud (Sugiyono, 2019). In this design, the treatment of the learning model consists of three levels (Standard Flipped Classroom / A1, Flipped Mastery Classroom / A2, and Conventional / A3), while the self-efficacy variable consists of two levels (High Self-Efficacy / B1 and Low Self-Efficacy / B2). The structure of the factorial design matrix 2 × 3 along with the symbolization of the measurement parameters is presented in detail in Table 1.

Table 1. ANCOVA Factorial Experiment Design Matrix 2 × 3

		Learning Model (A)		
		***Standard Flipped Classroom* (A1)**	***Flipped Mastery Classroom* (A2)**	**Convensional (A3)**
Self-Efficacy Level (B)	Height (B1)	x, μ_{A1B1}	x, μ_{A2B1}	x, μ_{A3B1}
	Low (B2)	x, μ_{A1B2}	x, μ_{A2B2}	x, μ_{A3B2}

Remarks: X = Covariate (Early Literacy and Numeracy Ability); μ = The average population of adjusted critical thinking ability for each treatment cell.

The target population in this study includes all students in class X of SMAN 1 Ubud for the 2025/2026 Academic Year which is evenly distributed into 9 parallel classes (X1 to X9) with a total population of 324 students. Each class is on average inhabited by 36 students with gender heterogeneity and institutionally homogeneous academic backgrounds.

In order to ensure that the populated classes were at an equal *baseline* before the intervention, a class equivalence test was carried out using One-Path Variance Analysis (ANOVA) on the initial pretest data on Informatics comprehension. The results of the ANOVA equivalence test between the 6 sample classes resulted in a value *of F = 1.087* with a p-value *of p = 0.369 > 0.05* (R-squared = 0.025), which statistically verified that there was no significant difference in initial ability between classes. Next, cluster *random sampling* assigns 3 sample classes:

- Class X-1 (N = 36): Treated as the *Standard Flipped Classroom* (A1) model.
- Class X-3 (N = 36): Was given the *Flipped Mastery Classroom* (A2) model treatment.
- Class X-5 (N = 36): Designated as a control class with *the Conventional* model (A3).

To form the moderator variable of self-efficacy (B), all students in the three sample classes (108 students) were measured their efficacy using a standard questionnaire. Based on the

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

extreme group design technique from Kelley (1939), 27% of students with the highest self-efficacy score were taken as the High Efficacy group (B1, n = 10 per class) and 27% of the students with the lowest score as the Low Efficacy group (B2, n = 10 per class). Thus, the final sample analyzed in factorial models 2 × 3 amounted to 60 students (30 students of the high efficacy group and 30 students of the low efficacy group who were spread evenly across 6 treatment cells).

Operationalization of Treatment Variables and Procedures

The variables in this study were operationally defined and controlled by a strict experimental procedure:

1. Independent Variables: Learning Model (A), consisting of:
   a) *Standard Flipped Classroom (A1):* Students access learning videos and e-modules in Google Classroom before class. The face-to-face session was filled with discussions on HOTS case studies, group LKS completion, and teacher clarifications.
   b) *Flipped Mastery Classroom (A2):* Students learn pre-class independent digital materials, followed by a multi-level formative completeness test (*mastery test* with *a passing grade* of 80%). Students who have completed progress to the advanced algorithm application project, while students who have not completed have received direct remediation before face-to-face. Classroom sessions are focused on synthesis and high-level consultation.
   c) *Conventional (A3):* Teacher-centered *learning* uses expository lecture methods, whiteboard demonstrations, individual classroom exercises, and traditional homework assignments.
2. Moderator Variable: Self-Efficacy Level (B), categorized into High Efficacy (B1) and Low Efficacy (B2) based on the distribution of the 27% upper and lower self-efficacy questionnaires.
3. Dependent Variables: Critical Thinking Ability (*Y*), measured post-treatment refers to the 6 dimensions Facione (2020): Interpretation, Analysis, Evaluation, Inference, Explanation, and Self-Regulation.
4. Control Variables (Covariate): Initial Literacy and Numeracy (*X*) ability, measured prior to treatment using an integrated diagnostic test that evaluates text comprehension, information extraction, and basic mathematical quantitative reasoning.

Empirical data collection uses three sets of standard instruments that have gone through the validity test stage *of expert judgment* and field empirical tests: The instrument is a reasoned multiple-choice objective test (5 options) of 50 items developed based on Facione's cognitive

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

indicators in the Data Analysis and Python Programming material. The content validity test with the Gregory formula with two evaluation experts resulted in a CV content validity coefficient = *1.00* (very high). The results of an empirical trial on 20 students using *Biserial Point* correlation (*r*) showed that 38 questions were consistently valid (*r > r = 0.444*), while 12 items were dropped. The reliability estimate for 38 valid questions was calculated using the Kuder-Richardson formula of 20 (KR-20) and obtained a very high reliability coefficient of *r = 0.954*. The reliability coefficient of 0.954 indicates that each question item in the instrument has a very strong internal consistency in measuring the same construct, namely the ability to think critically. In other words, there is a very high level of interconnectedness between the question items so that the instrument is able to produce consistent measurements when used in a group of respondents who have similar characteristics.

The self-efficacy questionnaire was developed based on a 5-point Likert scale covering 3 dimensions of Bandura: Magnitude (level of difficulty), Strength (strength of belief), and Generality (breadth of field). The initial draft amounted to 44 items. The results of the expert evaluation with the Gregory matrix eliminated 4 irrelevant items, resulting in 40 items with a validity of CV content = 1.00. The Pearson Product Moment correlation internal consistency test confirmed that 30 grains had an r> value of 0.444. Reliability tests using Alpha Cronbach hits yielded a fantastic value of α = 0.994 (excellent category).

The initial diagnostic test consisted of 40 multiple-choice questions based on the context of Informatics. Gregory's test results in a CV initial content validity index = 0.825. After discarding 7 irrelevant items, the 33 retained items were tested for item consistency with Point Biserial, where all 33 items were shown to be consistent (r ≥ 0.59 > 0.444). The reliability coefficient of the KR-20 was recorded as r = 0.982.

Data processing is carried out in stages including descriptive statistics and inferential statistics. Before hypothesis testing, a classical ANCOVA assumption test was performed:

1. **Residual Normality Test:** Using the Shapiro-Wilk and Kolmogorov-Smirnov methods (***α = 0.05***).
2. **Variance Homogeneity Test:** Using *Levene's Test of Equality of Error Variances* (***α = 0.05***).
3. **Homogeneity of Regression Slopes *Test*:** Tested the significance of the effect of the interaction between treatment factors with covariates (***A × B × X***). The assumption is tested if the p-value ***p > 0.05***.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

After all assumptions were met, the main hypothesis test was calculated using Two-Path Covariance Analysis (ANCOVA) at a significance level of ***α = 0.05*** assisted by the SPSS v.25 program. To see the detailed differences between cells, a *post-hoc test* based on *Pairwise Comparisons* with *Least Significant Difference* (LSD) adjustment was carried out.

RESULTS AND DISCUSSION

A summary of descriptive statistics for the covariate data of Initial Literacy & Numeracy Ability (***X***) and the bound variable Critical Thinking Ability (***Y***) in the 6 factorial groups is presented in Table 1.2 and Table 2.

Table 2. Statistical Description of Students' Initial Literacy & Numeracy Scores (Covariate X)

STATISTICAL SIZE	HIGH SELF-EFFICACY (B1)			LOW SELF-EFFICACY (B2)		
	SFC (A1B1)	FMC (A2B1)	CONV (A3B1)	SFC (A1B2)	FMC (A2B2)	CONV (A3B2)
Sample Count (N)	10	10	10	10	10	10
Mean	72.60	72.50	72.50	72.50	72.60	72.60
Median	72.50	72.50	72.50	72.50	72.50	72.50
Standard Deviation	4.55	4.74	4.35	4.35	4.47	4.40
Varians	20.71	22.50	18.94	18.94	20.04	19.38
Score Minimum	65.00	64.00	65.00	65.00	65.00	65.00
Score Maximum	81.00	81.00	80.00	80.00	80.00	80.00

The data in Table 2. show that the average initial literacy and numeracy (***X***) scores in all treatment cells were very identical and homogeneous, ranging from 72.50 to 72.60 with standard deviations ranging from 4.35–4.74. This confirms the equivalence of the subject's basic cognitive capacity before being given experimental treatment.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Table 3. Statistical Description of Post-Treatment Students' Critical Thinking Ability Score (Y-bound variable)

STATISTICAL SIZE	HIGH SELF-EFFICACY (B1)			LOW SELF-EFFICACY (B2)		
	SFC (A1B1)	FMC (A2B1)	CONV (A3B1)	SFC (A1B2)	FMC (A2B2)	ONV (A3B2)
Sample Count (N)	10	10	10	10	10	10
Mean	85.00	90.00	74.00	79.00	81.00	70.00
Median	85.00	90.00	74.00	79.00	81.00	70.00
Standard Deviation	2.21	2.21	2.21	2.21	2.21	2.21
Varians	4.89	4.89	4.89	4.89	4.89	4.89
Score Minimum	81.00	86.00	70.00	75.00	77.00	66.00
Score Maximum	89.00	94.00	78.00	83.00	85.00	74.00

Table 3. shows the variation in the pattern of achievement of critical thinking skills that are striking between groups. The highest average achievement was achieved by the group of students with high self-efficacy who were taught with *Flipped Mastery Classroom* (A2B1) of 90.00, followed by *the Standard Flipped Classroom* with high efficacy (A1B1) of 85.00. Meanwhile, the lowest achievement was experienced by the conventional learning group with low self-efficacy (A3B2) which only reached an average of 70.00.

Before the inferential hypothesis test, three prerequisite tests were strictly analyzed: The normality test of the distribution of the covariate (***X***) and bound (***Y***) score distributions was carried out with the Kolmogorov-Smirnov and Shapiro-Wilk tests on the 6 cell groups.

Table 4. Shapiro-Wilk Normality Test Results Covariate (X) and Bound (Y) Scores

CELL GROUPS	EARLY LITERACY & NUMERACY (X)			CRITICAL THINKING SKILLS (Y)		
	STATISTICS S-W	DF	SIG. (P)	STATISTICS S-W	DF	SIG. (P)
A1 B1	0.995	10	1.000	0.984	10	0.983

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

A2 B1	0.995	10	1.000	0.984	10	0.983
A3 B1	0.998	10	1.000	0.984	10	0.983
A1 B2	0.998	10	1.000	0.984	10	0.983
A2 B2	0.992	10	0.998	0.984	10	0.983
A3 B2	0.994	10	1.000	0.984	10	0.983

Based on Table 4, all Shapiro-Wilk significance values for variables ***X*** and ***Y*** in the 6 treatment cells had a p-value ***p > 0.05*** (p = 0.983 to 1,000). Thus, the data is perfectly distributed normally.

The similarity test of error variance between cells is calculated with *Levene's Test*. The statistical results obtained ***F(5, 54) = 1.584*** with p-value ***p = 0.180 > 0.05***. This confirms that the variance of the bound variable error is homogeneous across treatment groups.

Homogeneity of Slopes Test

The regression tilt test tested the absence of interaction between the covariate (***X***) and the non-moderator variable (***A × B***).

Table 5. Results of Evaluation of Regression Slope Homogeneity Assumptions (A × B × X)

SOURCE OF VARIATION	TYPE III SUM OF SQUARES	DF	MEAN SQUARE	F-COUNT	SIG. (P)
Corrected Model	2888.506	11	262.591	3293.682	0.000
Kovariat (X)	259.925	1	259.925	3260.236	0.000
A*B Interaction	12.006	5	2.401	30.119	0.000
Interaction A*B*X	0.229	5	0.046	0.575	**0.718**
Galat (Error)	3.827	48	0.080		

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

The test results in Table 5. show that the effect of 3-way interaction ***A × B × X*** results in a value ***of F = 0.575*** with a significance value of ***p = 0.718 > 0.05***. This means that the regression line of the covariate relationship between literacy and critical thinking has a parallel/homogeneous slope across all groups. Thus, all the requirements of linearity and the assumptions of ANCOVA's analysis are fulfilled absolutely.

Hypothesis testing was carried out by comparing the Two-Path ANCOVA (with X covariate control) and Two-Path ANOVA (without X covariate control) models.

Main Hypothesis Testing (Two-Track ANCOVA: Hypotheses 1, 3, and 5)

Table 6. presents a summary of the results of the ANCOVA Two-Path calculation to test the influence of the learning model (A), self-efficacy (B), and its interaction (A × B) by controlling for initial literacy-numeracy (X).

Table 6. Results of Covariance Analysis (ANCOVA) Two Pathways of Critical Thinking Ability

SOURCE OF VARIATION	TYPE III SUM OF SQUARES	DF	MEAN SQUARE	F-COUNT	SIG. (P)	PARTIAL H2
Corrected Model	2888.277	6	481.380	6289.848	0.000	0.999
Constant *(intercept)*	403.363	1	403.363	5270.458	0.000	0.990
Literacy Covariate-Numeracy (X)	259.944	1	259.944	3396.503	0.000	0.985
Learning Model (A)	1963.333	2	981.667	**12826.750**	**0.000**	0.998
Self-Efficacy (B)	604.762	1	604.762	**7901.998**	**0.000**	0.993
Interaction (A*B)	63.691	2	31.846	**416.106**	**0.000**	0.940
Galat (Error)	4.056	53	0.077			
Total	385294.000	60				

Based on the output of ANCOVA in Table 1.6, the hypothesis conclusions are formulated as follows:

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

1. Hypothesis Testing 1 (Influence of Learning Model by Controlling Covariates): F value = *12826.750* with *p value = 0.000 < 0.05*. Result: *H* is rejected and *H* is accepted. This means that *hit 0 a* there is a very significant difference/influence of the learning model on students' critical thinking skills after controlling for literacy and numeracy skills at SMAN 1 Ubud.
2. Hypothesis Testing 3 (Effect of Self-Efficacy by Controlling Covariates): F value = *7901.998* with *p value = 0.000 < 0.05*. Result: *H* is rejected and *H* is accepted. This means that there is *a very* significant influence of the level of self-efficacy on students' critical thinking skills after controlling for initial literacy-numeracy.
3. Hypothesis Testing 5 (Effect of Interaction A × B with Covariate Control): F value = *hit 416.106* with *p value = 0.000 < 0.05*. Result: *H* is rejected and *H* is accepted. This *0 a* proves that there is a significant interaction effect between the learning model and the level of self-efficacy on students' critical thinking skills after controlling for literacy and numeracy skills initially.

Comparison of Covariate-Controlled Tests (Two-Path ANOVA: Hypotheses 2, 4, and 6)

To evaluate the crucial role of the covariate variable X, parallel testing using Two-Path ANOVA (without literacy-numeracy control) is presented in Table 7.

Table 7. Two-Path Variance Analysis (ANOVA) Results Without Covariate Control

SOURCE OF VARIATION	TYPE III SUM OF SQUARES	DF	MEAN SQUARE	F-CALCULATE	SIG. (P)	PARTIAL H2
Corrected Model	2628.333	5	525.667	107.523	0.000	0.909
Constant *(intercept)*	382401.667	1	382401.667	78218.523	0.000	0.999
Learning Model (A)	1963.333	2	981.667	**200.795**	**0.000**	0.881
Self-Efficacy (B)	601.667	1	601.667	**123.068**	**0.000**	0.695
Interaction (A*B)	63.333	2	31.667	**6.477**	**0.003**	0.193
Galat (Error)	264.000	54	4.889			
Total	385294.000	60				

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

A comparison between Table 7 (ANCOVA) and Table 3 (ANOVA) shows a remarkable statistical contrast:

1. Without controlling for the covariate X, the Mean Square Error variance is relatively high at **4.889**. However, when the X covariate was included in the ANCOVA model, the error variance was drastically reduced by 98.4% to only **0.077**.
2. This decrease in error variance causes the F-count statistics for the Learning Model Key Effects to shoot from ***F = 200.795*** (ANOVA) to ***F = 12826.750*** (ANCOVA).
3. These findings scientifically prove that controlling students' initial literacy and numeracy skills is empirically proven to be able to eliminate innate cognitive noise, so that the sensitivity and accuracy of estimating the effect of *the Flipped Classroom* model's treatment are increased exponentially.

Simple and Post-Hoc Effects Analysis (Hypothesis 7 and Pairwise Comparisons)

To determine the specific differences between learning models in the group of students with high self-efficacy (B1) and controlling the X covariate, the ANCOVA One Special Pathway B1 test (Hypothesis 7) was carried out which is summarized in Table 8.

Table 8. ANCOVA Results One Path of Influence of Learning Model on High Self-Efficacy Group (B1)

SOURCE OF VARIATION	TYPE III SUM OF SQUARES	DF	MEAN SQUARE	F-COUNT	SIG. (P)
Corrected Model	1470.318	3	490.106	7576.848	0.000
Kovariat (X)	130.318	1	130.318	2014.668	0.000
Learning Model (A123B1)	1338.078	2	669.039	**10343.081**	**0.000**
Galat *(Error)*	1.682	26	0.065		
Total	208142.000	30			

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

International Journal of Education, Vocational and Social Science
https://e-journal.citakonsulfindo.or.id/index.php/IJEVSS
E-ISSN : 2963 4954
IJEVSS

Table 8. confirms the remarkable significance (***F = 10343.081; p = 0.000***), showing a sharp difference in critical thinking between models in high-efficacy students. Pairwise Comparisons with *adjusted marginal means* are presented in Table 9.

Table 9. *Post-Hoc Test Results of Pairwise Comparisons* Between Models in the High Efficacy Group (B1)

(I) MODEL (A)	(J) MODEL (A)	ADJUSTED AVERAGE DIFFERENCE (I - J)	STD.ERROR	SIG. (P)	95% CONFIDENCE INTERVAL
Standard Flipped (1)	*Flipped Mastery* (2) Conventional (3)	-5.048* +10.952*	0.114 0.114	0.000 0.000	[-5.282, -4.814] [10.718, 11.186]
Flipped Mastery (2)	Standard *Flipped* (1) Conventional (3)	+5.048* +16.000*	0.114 0.114	0.000 0.000	[4.814, 5.282] [15.766, 16.234]
Conventional (3)	Standard *Flipped (1)* *Flipped Mastery (2)*	-10.952* -16.000*	0.114 0.114	0.000 0.000	[-11.186, -10.718] [-16.234, -15.766]

The *adjusted mean* of critical thinking ability in the high-efficacy group was ordered hierarchically:

1. *Flipped Mastery Classroom* (A2B1): ***Mean = 90,016*** (Capaian tertinggi mutlak).
2. *Standard Flipped Classroom* (A1B1): ***Mean = 84,968.***
3. Konvensional (A3B1): ***Mean = 74,016.***

RESULTS AND DISCUSSION

The Superiority of Flipped Mastery Classroom in Triggering Critical Reasoning

The absolute advantage of the *Flipped Mastery Classroom* (FMC) model with an adjusted average score of 90,016 confirms the theory of cognitive constructivism and *mastery learning*. In

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Informatics learning, the material is strictly hierarchical; For example, understanding the structure of branching logic (*if-else*) and loops in Python requires a thorough mastery of variables and data types. In the FMC model, the existence of a formative completeness test gate (*mastery threshold*) before class ensures that no student carries "learning gaps" when entering the HOTS-based application phase in the classroom.

When the in-person session takes place, students who have completed 100% on the basic concepts can be directly involved in solving complex algorithm case study analysis, program deconstruction, and bug identification. This activity intensively stimulates Facione's critical thinking indicators, especially the Analysis, Evaluation, and Explanatory dimensions (Facione, 2020). In contrast, in conventional learning, lesson hours are sucked up for one-way lectures, so students never have the opportunity to explore high-level cognitive reasoning independently.

The Role of Self-Efficacy Moderation and the Construction of Self-Regulated Learning

Significant interaction findings (***F = 416.106; p < 0.001***) confirmed that self-efficacy was the psychological axis that determined the success of the *Flipped Classroom* model. Students with high self-efficacy have full confidence in their capacity to independently navigate pre-class e-modules and digital videos. They applied *Self-Regulated Learning* (SRL) strategies, such as managing study timing, noting critical points, and repeating videos on difficult parts (Lidiawati et al., 2021). Therefore, when entering the class discussion session, they are ready to dismantle scientific arguments and reflect on their thought process (*self-regulation*).

Conversely, for students with low self-efficacy, the absence of direct teacher assistance when studying pre-class digital materials can cause anxiety and cognitive barriers. However, the *Flipped Mastery Classroom* model has been shown to be able to mitigate the weaknesses of low-efficacy students gradually through individual remediation schemes and immediate *feedback*. The experience of successfully completing the *mastery experiences* provides new ammunition for strengthening students' self-efficacy, which in turn boosts their critical thinking achievements beyond conventional classes (Lestari, 2024).

The Urgency of Controlling Covariates of Literacy and Early Numeracy

Methodologically, the success of controlling the initial literacy and numeracy covariates in ANCOVA which was able to cut the error variance by up to 98.4% made a real contribution to the educational research methodology in Indonesia. Reading literacy is a means of decoding

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

conceptual information, while numeracy is a means of systematic logical reasoning. In the subject of Informatics, these two initial abilities work as basic cognitive processing machines. By neutralizing the inherent differences in students' initial literacy-numeracy, the results of this study provide valid and unbiased evidence that the intervention of the *Flipped Mastery Classroom* model is indeed purely the main cause of improving students' critical thinking skills.

CONCLUSION

Based on the results of the inferential data analysis and academic discussions that have been presented, several main conclusions can be drawn as follows: 1) There are very significant differences and influences of the *Flipped Classroom* learning model on students' critical thinking skills in Informatics subjects at SMAN 1 Ubud after statistically controlling the initial literacy and numeracy skills (***F = 12826.750; p < 0.001***). 2) There was a very significant influence of the level of self-efficacy on students' critical thinking skills after controlling for initial literacy and numeracy skills (***F = 7901.998; p < 0.001***), where students with high self-efficacy consistently outperformed students with low self-efficacy. 3) There was a significant effect of interaction between the learning model and the level of self-efficacy on students' critical thinking skills after controlling for initial literacy-numeracy (***F = 416.106; p < 0.001***), which suggests that the effectiveness of the learning model depends on the level of student self-confidence. 4) *The Flipped Mastery Classroom* (FMC) model is proven to be the most superior and effective model in boosting students' critical thinking skills, both in the high self-efficacy (***Adjusted Mean = 90,016***) and low self-efficacy groups, compared to *the Standard Flipped Classroom* and conventional models. 5) Control of literacy and numeracy covariate variables is proven to methodologically improve precision, internal validity, and ANCOVA's statistical test power by reducing the variance of experimental errors by up to 98.4%.

REFERENCES

Amalia, R., Nurfiyani, N., & Rahmawati, I. (2023). Integrasi Blended Learning dan Flipped Classroom dalam Pembelajaran Abad 21. *Jurnal Teknologi Pendidikan dan Kurikulum*, 11(2), 145–158.

Ariyanto, A., Suparman, S., & Kusumah, Y. S. (2020). Improving Higher-Order Thinking Skills via Problem-Based Learning in Digital Era. *International Journal of Instruction*, 13(4), 201–216.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

Aryanti, D. (2020). Peran Self-Efficacy terhadap Kemandirian Belajar dan Pemecahan Masalah Siswa. *Jurnal Psikologi Pendidikan & Konseling*, 6(1), 34–42.

Assessment in Education: Principles, Policy & Practice, 25(1), 5–20.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.

Cipta, A., Kurniawan, R., & Sori, M. (2024). Analisis Kendala Pembelajaran Konvensional terhadap Penalaran Kritis Siswa SMA. *Jurnal Inovasi Pendidikan Menengah*, 5(1), 88–97.

Dwijayanti, P., Gunawan, I., & Suastra, I. W. (2022). Pengembangan Instrumen Evaluasi Keterampilan Berpikir Kritis Berbasis High-Order Thinking Skills. *Jurnal Penelitian dan Evaluasi Pendidikan*, 26(2), 112–125.

Ekayana, I. N. (2024). *Pengembangan Skala Efikasi Diri Siswa SMA dalam Pembelajaran Sains Terintegrasi*. Undiksha Press.

Erlin, E., Dahlan, J. A., & Juandi, D. (2021). The Effectiveness of 4C-Based Learning Framework in Enhancing Critical Thinking Skills: A Meta-Analysis. *Journal of Educational Sciences*, 15(3), 301–315.

Facione, P. A. (2020). *Critical Thinking: What It Is and Why It Counts* (2020 ed.). Measured Reasons LLC.

Fitriani, N., Suhendra, S., & Turmudi, T. (2022). Implementation of Flipped Classroom Model to Enhance Students' Mathematical Critical Thinking. *Journal of Physics: Conference Series*, 2155(1), 012014.

Hasanah, U., & Lusita, R. (2023). Transformasi Kurikulum Merdeka dalam Mendorong Student-Centered Learning di Sekolah Menengah. *Jurnal Kebijakan dan Pengembangan Pendidikan*, 11(1), 45–59.

Herlina, L., Rahayu, S., & Sugiarto, B. (2020). *Metodologi Penelitian Pendidikan Kuantitatif dan Eksperimen*. Rajawali Pers.

Lestari, W. (2024). Efektivitas Model Flipped Mastery Classroom dan Strategi Backtracking terhadap Self-Efficacy dan Hasil Belajar. *Jurnal Riset Pendidikan Matematika*, 11(1), 76–89.

Lidiawati, K. R., Susanti, E., & Setiawan, A. (2021). Self-Efficacy and Academic Self-Regulation in Online Learning Environment. *Indonesian Journal of Educational Psychology*, 2(2), 99–110.

Nainggolan, H., & Panjaitan, M. (2025). The Impact of Flipped Classroom on Students' Critical Thinking Skills and Learning Persistence. *Asian Journal of Educational Technology*, 8(1), 12–27.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

OECD. (2023). *PISA 2022 Results (Volume I): The State of Learning and Equity in Education*. OECD Publishing. https://doi.org/10.1787/53f238b3-en

Sälzer, C., & Roczen, N. (2018). Assessing Global Competence in PISA 2018: Challenges and Opportunities.

Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.

.

Copyright at authors some right reserved this work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License