JTAM (Jurnal Teori dan Aplikasi Matematika)
http://journal.ummat.ac.id/index.php/jtam
p-ISSN 2597-7512 | e-ISSN 2614-1175
Vol. 10, No. 2, April 2026, pp. 442-452

Fuzzy C-Means Clustering of Student Mathematical Communication Skills and Cognitive Performance and Its Association with Learning Models

Dhidhi Pambudi[1*], Sachiko Nakano[2], Alfina Damayanti[1]
[1]Dept. of Mathematics Education, Sebelas Maret University, Indonesia
[2]Graduate School of Humanities and Economic Sciences, Yamaguchi University, Japan
dhidhipambudi@staff.uns.ac.id

Article History:
Received : 09-10-2025
Revised : 17-12-2025
Accepted : 31-12-2025
Online : 01-04-2026

Keywords:
Educational Data Mining;
Fuzzy C-Means;
Learning Cycle 7E;
Mathematical Communication;
Cognitive Style.

ABSTRACT

This study applies Educational Data Mining (EDM) to examine how students' mathematical communication skills and cognitive styles interact under different instructional contexts. A quantitative research that applies quasi-experimental design combined with EDM was employed involving 64 seventh-grade students at SMP Negeri 2 Jaten, divided into Learning Cycle 7E (LC7e) and Direct Instruction (DI) groups. The research instruments included a validated essay test for mathematical communication and the Group Embedded Figures Test (GEFT). Data were analyzed using the Fuzzy C-Means (FCM) algorithm to partition students into distinct clusters based on their cognitive-communicative attributes. The analysis resulted in a stable five-cluster model, representing progressive levels of cognitive–communicative integration. The model shows that cognitive profile predominantly creates cluster structure, whereas the Learning Cycle 7E (LC7E) model exerts a moderating influence. Students taught through LC7E were more concentrated in higher-performing clusters than those in DI classrooms. Furthermore, Field-Independent (FI) learners tended to achieve the highest communicative profiles, yet Field-Dependent (FD) learners also benefited meaningfully from LC7E activities that emphasized exploration and reflection. These results demonstrate that the LC7E model supports cognitive and communicative development across the learner spectrum, with differentiated gains linked to cognitive style. These findings highlight the utility of EDM in capturing student heterogeneity and provide a basis for educators to design adaptive learning strategies that accommodate diverse cognitive characteristics.

https://doi.org/10.31764/jtam.v10i2.35522

This is an open access article under the CC–BY-SA license

A. INTRODUCTION

The rapid integration of digital technologies into education has generated large volumes of student learning data, creating opportunities for Educational Data Mining (EDM) to uncover hidden patterns and support evidence-based decision making (Khine, 2024; Romero & Ventura, 2020). For decades, educational research has predominantly relied on traditional methods rooted in classical inferential statistics, such as Analysis of Variance (ANOVA), which focus primarily on mean differences between pre-defined groups. While useful, these mean-based approaches often overlook the heterogeneity within student populations, effectively masking granular individual differences. To address this critical gap, EDM applies machine learning and data analytics methods to educational datasets to discover latent structures, predict outcomes, and inform personalized interventions (Baker & Siemens, 2014). Among machine learning

methods, clustering is a fundamental unsupervised learning approach for identifying natural groupings of students without predefined labels (Han et al., 2012). Unlike aggregate statistical comparisons, clustering reveals distinct profiles of learners who share similar cognitive or behavioral characteristics, thereby enabling instructors to design more precise, adaptive pedagogical strategies (Papamitsiou & Economides, 2014).

In mathematics education, mathematical communication skill the ability to express mathematical ideas coherently, logically, and accurately is a core competence linked to conceptual understanding and problem solving (Hwang et al., 2007; NCTM, 2000). International assessments such as PISA continue to show that Indonesian students struggle with tasks requiring reasoning and communication (OECD, 2019). Research has also highlighted the importance of cognitive style, particularly the distinction between Field-Independent (FI) and Field-Dependent (FD) learners (Witkin et al., 1977). FI students tend to analyze information analytically and work autonomously, whereas FD students process information more holistically and depend on external cues. Empirical studies indicate that FI learners generally achieve higher performance in tasks that demand structured reasoning and symbolic representation (Tinajero & Páramo, 1997; Zhang & Sternberg, 2005). Foundational research by Witkin and colleagues established that FI individuals exhibit greater perceptual disembedding ability, which supports independent reasoning and analytical problem solving (Witkin et al., 1977). Subsequent studies have shown that FD learners, while sometimes disadvantaged in abstract tasks, often benefit from instructional support, contextualized explanations, and collaborative learning environments (Riding & Cheema, 1991). Furthermore, cognitive style should not be interpreted as a measure of ability, but rather as a preference in information processing that interacts with instructional design (Messick, 1984). Together, these findings suggest that cognitive style plays a meaningful role in shaping how students engage with mathematical reasoning and communication tasks, reinforcing the importance of instructional approaches that accommodate cognitive diversity. Understanding this communication-cognitive diversity is critical for designing adaptive instructional models, as it further impacts learning outcomes.

The Learning Cycle 7E (LC7E) model, an extension of the 5E Bybee et al. (2006) inquiry cycle, engages students through sequential phases of elicit, engage, explore, explain, elaborate, evaluate, and extend (Eisenkraft, 2003). LC7E emphasizes iterative exploration and reflection, which have been shown to improve students' conceptual understanding and communication of scientific and mathematical ideas (Nurochmah et al., 2021). However, most previous studies have examined these relationships using classical inferential statistics such as ANOVA, which test group differences but offer limited insight into the latent structure of student populations. A critical limitation of these traditional methods is their tendency to impose rigid, pre-defined grouping on student populations. By forcing data into fixed experimental categories, ANOVA often presents an idealized view of learning outcomes, effectively masking naturally occurring latent subgroups. This artificial rigidity fails to capture the granular heterogeneity of students, potentially leading to generalized conclusions that overlook the complex interplay between cognitive styles and communication skills. To address this gap, the present study applies Fuzzy C-Means (FCM) clustering Bezdek et al. (1984) to identify natural groupings of students based on continuous scores of mathematical communication skill and cognitive performance

measured by the Group Embedded Figures Test (GEFT). Unlike hard-partitioning algorithms such as K-Means, FCM assigns membership degrees to each cluster, capturing the overlapping characteristics typical of educational data (Gan et al., 2007). The cluster obtained can be examined by considering its association with learning models (LC7E vs. Direct Instruction) to reveal how instructional context aligns with emergent communication–cognitive profiles.

This study integrates EDM techniques with mathematics education theory to uncover latent learning patterns. Accordingly, this study aims to: (i) identify natural groupings of students based on their cognitive performance and mathematical communication skills using Fuzzy C-Means clustering, and (ii) examine the relationship of these clusters with learning models regarding the cognitive style within the framework of EDM. The findings are expected to inform adaptive learning strategies and support the development of intelligent tutoring systems capable of personalizing instruction according to students' cognitive and communicative characteristics.

B. METHODS

1. Participants and Setting

The study was a quantitative research that applied a quasi-experimental design combined with EDM conducted at SMP Negeri 2 Jaten, a state junior high school located in Karanganyar Regency, Central Java, Indonesia. The participants were seventh-grade students enrolled in mathematics classes during the 2023 academic year. A total of 64 students (approximately balanced by gender) participated. Two intact classes were used: one received instruction using the Learning Cycle 7E (LC7E) model, and the other followed a Direct Instruction (DI) model, following the quasi-experimental design described by Capili & Anastasi (2024). Prior to the intervention, students' cognitive styles were assessed using the Group Embedded Figures Test (GEFT) to classify their cognitive profiles. Additionally, to ensure comparability, statistical analysis confirmed that both groups were at a statistically equivalent baseline regarding their initial mathematical communication skills prior to the intervention. All students and school administrators provided consent consistent with local ethical guidelines.

2. Instruments

a. Mathematical Communication Test

Students' mathematical communication ability was assessed using a researcher-developed test consisting of open-ended items that required explanation of reasoning, representation with mathematical symbols, and verbal interpretation of solutions (NCTM, 2000). The test has five cases that underwent content validation by subject matter experts. Psychometric analysis demonstrated that the items possessed good discriminatory power, with coefficients ranging from 0.737 to 0.889, and a moderate difficulty level (indices between 0.535 and 0.75). Furthermore, the instrument exhibited high reliability with a coefficient of 0.905. The final test produced a composite score in the range of 0–27.

b. Group Embedded Figures Test (GEFT)
Cognitive style was measured using the Group Embedded Figures Test (GEFT). This is a standard test, firstly developed by Witkin et al. and widely used in psychology-education research globally. The GEFT asks students to identify simple shapes embedded in complex figures (Witkin et al., 1977). The GEFT yields a performance score in the range 0–18, where higher values indicate stronger Field-Independent (FI) tendencies and lower values indicate Field-Dependent (FD) tendencies. Although FI/FD categories can be derived by thresholding (Khatib & Hosseinpur, 2011), this present study applies the raw GEFT scores to retain metric information for clustering.

c. Data Preparation
Both continuous features (mathematical communication and GEFT score) were inspected for completeness and standardized (z-scores) to remove scale effects (Han et al., 2012). No missing data were detected. The learning model (LC7E vs. DI) was retained as a nominal variable for post-hoc interpretation but was not included in the clustering procedure.

d. Clustering Procedure
Clustering was performed using the Fuzzy C-Means (FCM) algorithm (Bezdek et al., 1984), a soft-partitioning method that assigns each student a membership degree for every cluster. The algorithm minimizes the objective function

$$J_m = \sum_{i=1}^{N} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - v_j \right\|^2 \quad (1)$$

where u_{ij} is the membership of student i in cluster j, v_j is the cluster center, and m is the fuzziness parameter (set to m = 2). Parameters were initially set to a maximum of 500 iterations, with the random seed adjusted iteratively to obtain a stable clustering configuration. Stability here refers to the consistency of the resulting partitions when evaluated under both information-theoretic criteria the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) (Zhang et al., 2023), which often yield slightly divergent optimal solutions. The final seed value was selected when both AIC and BIC converged to the same clustering structure, indicating that the model achieved a balance between goodness of fit and parsimony. This approach minimizes random initialization bias and enhances the reproducibility of the clustering outcome (Burnham & Anderson, 2004; Claeskens & Hjort, 2001).

e. Statistical Analysis of Associations
After clustering, each student was assigned to the cluster corresponding to their highest membership value. The relationship between cluster membership and the learning model (LC7E vs. DI) was evaluated using a Chi-square test of independence. A contingency table was generated to describe the frequency distribution of students across clusters and instructional models. Expected cell counts were examined to confirm that Chi-square assumptions were satisfied. To complement this test, Cramér's V was computed to assess the strength of association between the categorical features (Cramér, 1946; Kim, 2017). All computations, including the clustering, were performed using open-source statistical software JASP (JASP Team, 2025).

C. RESULT AND DISCUSSION

1. Cluster Result

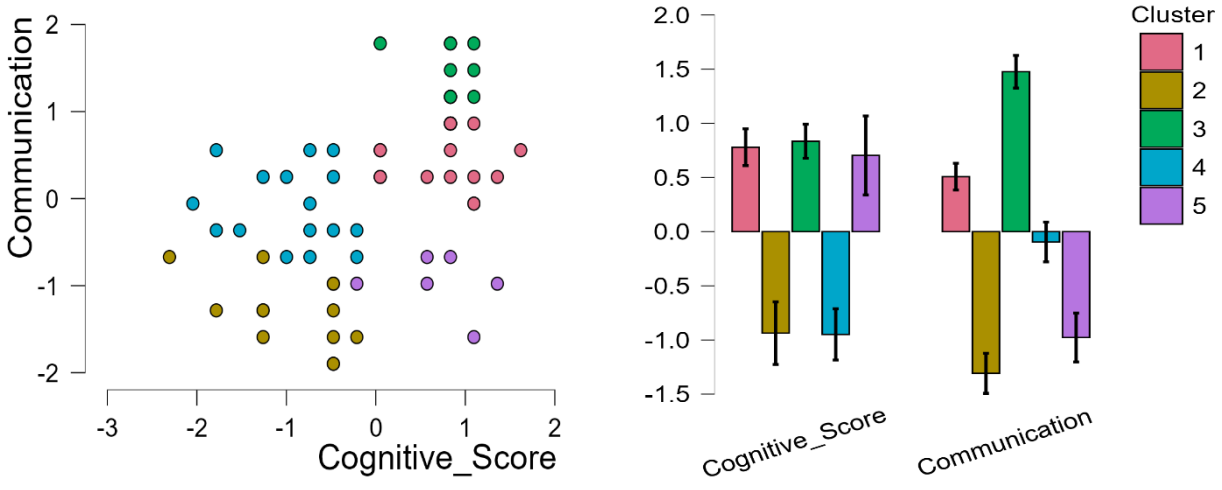
The clustering procedure employed the Fuzzy C-Means (FCM) algorithm with fuzziness parameter m=2, a maximum of 500 iterations, and an initial maximum of ten clusters. After iterative testing, the random seed was fixed at 5, yielding a stable clustering configuration under both the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). The resulting model (AIC = 43.92; BIC = 65.50; R^2=0.802) showed strong model fit, and crucially, both AIC and BIC converged to the same solution, confirming the internal consistency and reliability of the partition (Burnham & Anderson, 2004; Zhang et al., 2023). Those results indicate that the model achieved a stable configuration balancing fit and parsimony (Singh & Singh, 2024; Wani, 2024). Five clusters (C1–C5) were identified, as summarized in Table 1 and visualized in Figure 1. The descriptive statistics are provided in Table 2. Each cluster represents a distinct performance profile derived from students' mathematical communication skills and cognitive performance.

Table 1. Cluster information

Cluster	**1**	**2**	**3**	**4**	**5**
Size	19	13	10	16	6
Explained proportion within-cluster heterogeneity	0.233	0.284	0.066	0.332	0.085
Within sum of squares	5.572	6.784	1.574	7.948	2.037

Table 2. Descriptive statistics

	Cognitive					**Math Communication**				
Cluster	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
Valid	19	13	10	16	6	19	13	10	16	6
Missing	0	0	0	0	0	0	0	0	0	0
Mean	11.79	5.231	12	5.188	11.5	21.84	15.92	25	19.88	17
Stdev	1.718	2.421	1.155	2.198	2.074	1.068	1.32	0.943	1.455	1.095
Minimum	9	0	9	1	8	20	14	24	18	15
Maximum	15	8	13	8	14	23	18	26	22	18



(a) Cluster matrix plot

(b) Cluster mean plot

Figure 1. Cluster plots

Cluster description:

a. C1, cluster 1 (High Cognitive – High Communication):
   Students in this cluster obtained above-average GEFT scores (mean = 11.79) and above-average communication scores (mean = 21.84), indicating comparatively stronger analytical performance and clearer mathematical expression in the test items.
b. C2, cluster 2 (Low Cognitive – Low Communication):
   This cluster had the lowest mean scores for both dimensions, representing students who struggled with disembedding complex visual information and expressing reasoning in mathematical form.
c. C3, cluster 3 (Very High Cognitive – Very High Communication):
   This is the top-performing cluster that features the highest cognitive scores and communication performance. Students here exemplify optimal integration between analytical reasoning and verbal or symbolic expression.
d. C4, cluster 4 (Low Cognitive – Moderate Communication):
   Members of this cluster displayed limited abstract reasoning ability but moderate communication proficiency.
e. C5, cluster 5 (Moderate Cognitive – Moderate-High Communication):
   An intermediate cluster bridging low and high performers. The students achieved moderate GEFT scores and moderately high communication performance, implying a transitional zone between cognitive and communicative development.

This pattern reflects the heterogeneity of cognitive and communicative profiles in middle school mathematics. The presence of moderate-overlap clusters (4 and 5) supports the fuzzy assumption that learning-related behaviors exist on a continuum rather than in discrete categories, which models gradual transitions among learner characteristics (Kalita et al., 2025). This is an important insight that fuzzy clustering naturally captures (Pérez-Ortega et al., 2024)

2. Association Between Clusters and Learning Models

The association between the five identified clusters and the two learning models (LC7E and DI) was examined using the Chi-square test of independence. The descriptive distribution is presented in Table 3, and inferential results appear in Table 4. The result indicates no statistically significant association between cluster membership and learning model at the 0.05 level.

Table 3. Contingency (Cluster vs Learning Model)

Cluster	Characteristic		Model		
	Cognitive	Math. Comm.	DI	LC7E	Total
C1	High	High	7	12	19
C2	Low	Low	9	4	13
C3	Very High	Very High	3	7	10
C4	Low	Moderate	9	7	16
C5	Moderate	Moderate-High	4	2	6
		Total	32	32	64

Table 4. Chi-Squared (Cluster vs Learning Model)

	Value	df	p
c^2	5.756	4	0.218
Likelihood ratio	5.881	4	0.208
Contingency coefficient	0.287		
Cramer's V	0.3		

Although no significant association, the pattern shows that learners taught with the LC7E model were more concentrated in Clusters 1 (12 out of 19, 63.16%) and Cluster 3 (7 out of 10, 70%), both characterized by high cognitive performance and strong mathematical communication. In contrast, students in the DI group were more frequent in Clusters 2 (9 out of 13, 69.23%) and Cluster 4 (9 out 16, 56.25%), where both cognitive and communication levels were lower. The pattern confirms that the directional tendency is pedagogically meaningful. It suggests that the LC7E model may better support the integration of analytical reasoning and expressive skill, consistent with prior evidence that active and inquiry-based cycles promote deeper conceptual understanding (Kalita et al., 2025).

Based on the values of Cramér's V = 0.300 and the Contingency Coefficient = 0.287, those reflect a weak-to-moderate relationship, reinforcing the conclusion that while not statistically strong, the LC7E model exerted a measurable influence on the clustering structure. Such modest yet systematic effects are common in educational data mining, where learning behaviors interact with cognitive characteristics in nonlinear ways (Singh & Singh, 2024). Those findings illustrate the usefulness of unsupervised approaches such as FCM for uncovering latent learner groupings that may not appear in traditional hypothesis testing. Even small directional effects can highlight meaningful instructional differentiation deserving of further investigation (Wani, 2024).

3. Cluster Interpretation and Impact of the Learning Model Regarding Cognitive Style

Starting from actual cognitive performance scores (GEFT), this study proceeds to the categorical cognitive styles: the FI and FD. As shown in Table 5, clusters with higher cognitive performance are occupied mainly by FI learners. The FI learners were concentrated in C1 (73.68%) and C3 (90%), whereas the FD learners appeared more often in C2 (100%) and C4 (100%). The Chi-square test, as shown in Table 6, revealed a significant association between cluster membership and cognitive style, indicating a strong relationship between clusters and cognitive tendencies. The pattern supports previous findings that FI learners are typically self-regulated in problem solving, enabling stronger performance in reasoning-oriented tasks (Nozari & Siamian, 2015). Notably, LC7E benefited both FI and FD learners, though to different extents. FI students achieved the greatest cognitive and communicative gains, while FD students also showed measurable improvement through the model's structured, supportive learning process.

Table 5. Contingency (Cluster vs Cognitive Style)

Cluster	Characteristic		Cog. Style		
	Cognitive	**Math. Comm.**	**FD**	**FI**	**Total**
C1	High	High	5	14	19
C2	Low	Low	13	0	13
C3	Very High	Very High	1	9	10
C4	Low	Moderate	16	0	16
C5	Moderate	Moderate-High	3	3	6
		Total	38	26	64

Table 6. Chi-Squared (Cluster vs Cognitive Style)

	Value	**df**	**p**
c^2	38.78	4	< 0.001
Likelihood ratio	49.74	4	< 0.001
Contingency coefficient	0.614		
Cramer's V	0.778		

These results imply that the effectiveness of LC7E is distributed along a continuum of cognitive ability rather than restricted to a single learner type. From a methodological perspective, the fuzzy clustering approach enabled detection of such graded relationships, which traditional mean-based methods might obscure. This reinforces the value of combining EDM and Learning Analytics (LA) with pedagogical analysis to understand how instructional models interact with cognitive diversity (Barbeiro et al., 2024; Kalita et al., 2025).

4. Synthesis

The clustering analysis revealed that cognitive performance plays a dominant role in shaping students' mathematical communication skills, while the learning model acts as a supporting influence. Through the Fuzzy C-Means method, a clear gradient of student profiles emerged from low to high levels of combined cognitive and communicative ability indicating that student performance varies along a continuum rather than forming discrete categories. This reinforces the usefulness of educational data mining approaches, which can detect subtle performance structures that may not appear through conventional analysis (Li et al., 2019; Romero & Ventura, 2020).

In general, students taught using the Learning Cycle 7E (LC7E) model were more represented in clusters with higher cognitive and communication scores, suggesting that this approach has a positive association with improved learning outcomes. The pattern also indicates that LC7E tends to be most beneficial for students with higher cognitive performance, though those with lower scores also show measurable impact. This finding aligns with evidence that structured exploration and reflection can enhance performance across a broad range of student abilities (Siantuba et al., 2023).

Consistent with prior studies Jati et al. (2017); Rahmy et al. (2020); Risdiana Chandra Dhewy (2022), our results confirm the superiority of the LC7E model in enhancing mathematical communication skills. However, this study transcends the limitations of traditional comparative statistics (ANOVA), which often mask individual heterogeneity behind aggregate means. By integrating Fuzzy C-Means clustering within an educational data mining framework, we provide distinct empirical evidence that learning effectiveness exists on a

continuous spectrum and that the instructional benefits of LC7E are robust across natural student groupings. This approach yields a more detailed picture of how cognitive and communicative outcomes interact with instructional methods, offering a nuanced validation of the model that goes beyond average scores. Ultimately, these findings provide educators with data-informed insights for differentiating instruction and understanding learner variability more precisely.

D. CONCLUSION AND SUGGESTIONS

This study employed Educational Data Mining (EDM) through Fuzzy C-Means (FCM) clustering to analyze patterns of students' cognitive performance and mathematical communication skills, and to examine their association with two instructional models: Learning Cycle 7E (LC7E) and Direct Instruction (DI) regarding the cognitive style. The clustering process identified five distinct student groups: C1 and C3 are higher cognitive-communication clusters; while C2, C3, and C5 are lower-moderate cognitive-communication clusters. The results revealed that cognitive performance serves as the principal factor shaping students' mathematical communication abilities, while the learning model functions as a moderating influence. Students engaged in LC7E were more frequently distributed among higher-performing clusters, which is 63.16% in C1 and 70% in C3, indicating that this model contributes positively to both cognitive and communicative development. Although the most pronounced improvements appeared among students with higher cognitive performance, learners with lower performance levels also experienced measurable gains. This finding suggests that LC7E supports a gradual continuum of learning benefits, emphasizing inclusivity across diverse cognitive profiles. These outcomes demonstrate the capability of EDM to reveal continuous performance structures often masked in traditional inferential analyses, offering deeper insights into how instructional methods interact with learner variability. Pedagogically, these findings advocate for adaptive inquiry models like LC7E that accommodate distinct cognitive profiles (FI/FD). While integrating such clustering analytics into digital platforms offers promise for real-time personalization, future research must address current limitations by expanding the sample size and incorporating affective or behavioral indicators to enhance model robustness and external validity.

ACKNOWLEDGEMENT

Sincere thanks to the administration, teachers, and students of SMP Negeri 2 Jaten, Karanganyar, Central Java, for their active participation and support during the data collection process. This research was funded by Sebelas Maret University, through the PKGR A Grant Program, Contract No. 371/UN27.22/PT.01.03/2025.

REFERENCES

Baker, R., & Siemens, G. (2014). Educational Data Mining and Learning Analytics. In *The Cambridge Handbook of the Learning Sciences* (pp. 253–272). Cambridge University Press. https://doi.org/10.1017/CBO9781139519526.016

Barbeiro, L., Gomes, A., Correia, F. B., & Bernardino, J. (2024). A Review of Educational Data Mining Trends. *Procedia Computer Science*, *237*, 88–95. https://doi.org/10.1016/j.procs.2024.05.083

Bezdek, J. C., Ehrlich, R., & Full, W. (1984). FCM: The fuzzy c-means clustering algorithm. *Computers & Geosciences*, *10*(2–3), 191–203. https://doi.org/10.1016/0098-3004(84)90020-7

Burnham, K. P., & Anderson, D. R. (Eds.). (2004). *Model Selection and Multimodel Inference*. Springer New York. https://doi.org/10.1007/b97636

Bybee, R. W., Taylor, J. a, Gardner, a, Scotter, P. V, Powell, J. C., Westbrook, a, & Landes, N. (2006). The BSCS 5E Instructional Model: Origins, Effectiveness, and Applications. In *BSCS*. https://bscs.org/wp-content/uploads/2022/01/bscs_5e_executive_summary-1.pdf

Capili, B., & Anastasi, J. K. (2024). An Introduction to Types of Quasi-Experimental Designs. *AJN, American Journal of Nursing*, *124*(11), 50–52. https://doi.org/10.1097/01.NAJ.0001081740.74815.20

Claeskens, G., & Hjort, N. L. (2001). *Model Selection and Model Averaging*. Cambridge University Press. https://doi.org/10.1017/CBO9780511790485

Cramér, H. (1946). *Mathematical Methods of Statistics*. Princeton University Press. https://archive.org/details/in.ernet.dli.2015.223699

Eisenkraft, A. (2003). Expanding the 5E Model: A proposed 7E model emphasizes "transfer of learning" and the importance of eliciting prior understanding. *The Science Teacher*, *70*(6), 56–59. http://www.jstor.org/stable/24156091

Gan, G., Ma, C., & Wu, J. (2007). *Data Clustering: Theory, Algorithms, and Applications*. Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9780898718348

Han, Jiawei., Kamber, Micheline., & Pei, Jian. (2012). *Data mining : concepts and techniques*. Elsevier/Morgan Kaufmann.

Hwang, W.-Y., Chen, N.-S., Dung, J.-J., & Yang, Y.-L. (2007). Multiple Representation Skills and Creativity Effects on Mathematical Problem Solving using a Multimedia Whiteboard System. *Educational Technology & Society*, *10*(2), 191–212. https://www.researchgate.net/publication/316239984

JASP Team. (2025). *JASP (Version 0.95.3)[Computer software]*. https://jasp-stats.org/

Jati, N. H. D., Budiyono, & Slamet, I. (2017). Students' Mathematical Communication Ability using Learning Cycle 7E on Junior High School. *Journal of Physics: Conference Series*, *895*, 012040. https://doi.org/10.1088/1742-6596/895/1/012040

Kalita, E., Oyelere, S. S., Gaftandzhieva, S., Rajesh, K. N. V. P. S., Jagatheesaperumal, S. K., Mohamed, A., Elbarawy, Y. M., Desuky, A. S., Hussain, S., Cifci, M. A., Theodorou, P., Hilčenko, S., Hazarika, J., & Ali, T. (2025). Educational data mining: a 10-year review. *Discover Computing*, *28*(1), 81. https://doi.org/10.1007/s10791-025-09589-z

Khatib, M., & Hosseinpur, R. M. (2011). On the Validity of the Group Embedded Figure Test (GEFT). *Journal of Language Teaching and Research*, *2*(3), 640–648. https://doi.org/10.4304/jltr.2.3.640-648

Khine, M. S. (2024). Educational Data Mining and Learning Analytics. In *Artificial Intelligence in Education* (pp. 1–159). Springer Nature Singapore. https://doi.org/10.1007/978-981-97-9350-1_1

Kim, H.-Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, *42*(2), 152. https://doi.org/10.5395/rde.2017.42.2.152

Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist*, *19*(2), 59–74. https://doi.org/10.1080/00461528409529283

NCTM. (2000). *Principles and Standards for School Mathematics*. National Council of Teachers of Mathematics. https://www.nctm.org/standards-and-positions/principles-and-standards/

Nozari, A., & Siamian, H. (2015). The Relationship between Field Dependent-Independent Cognitive Style and Understanding of English Text Reading and Academic Success. *Materia Socio Medica*, *27*(1), 39. https://doi.org/10.5455/msm.2014.27.39-41

Nurochmah, Arifin, S., & Surgandini, A. (2021). The application of learning cycle 7E model to improve a mathematics communication skills of junior high school students. *AIP Conference Proceedings 2331*, 020001. https://doi.org/10.1063/5.0041644

OECD. (2019). *PISA 2018 Results (Volume I)*. OECD. https://doi.org/10.1787/5f07c754-en

Papamitsiou, Z., & Economides, A. A. (2014). Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence. *Educational Technology & Society*, *17*(4), 49–64. https://www.jstor.org/stable/jeductechsoci.17.4.49

Pérez-Ortega, J., Moreno-Calderón, C. F., Roblero-Aguilar, S. S., Almanza-Ortega, N. N., Frausto-Solís, J., Pazos-Rangel, R., & Martínez-Rebollar, A. (2024). Hybrid Fuzzy C-Means Clustering Algorithm, Improving Solution Quality and Reducing Computational Complexity. *Axioms*, *13*(9), 592. https://doi.org/10.3390/axioms13090592

Rahmy, S. N., Usodo, B., & Slamet, I. (2020). Students' mathematical communication ability using 7E learning cycle based on students thinking style. *Journal of Physics: Conference Series*, *1469*(1), 012154. https://doi.org/10.1088/1742-6596/1469/1/012154

Riding, R., & Cheema, I. (1991). Cognitive Styles—an overview and integration. *Educational Psychology*, *11*(3–4), 193–215. https://doi.org/10.1080/0144341910110301

Risdiana Chandra Dhewy. (2022). Pengaruh Learning Cycle 7E Terhadap Kemampuan Komunikasi Matematis Peserta didik. *Jurnal Aplikasi Matematika Dan Statistik*, *1*(1), 21–26. https://doi.org/10.53625/jams.v1i1.3944

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *WIREs Data Mining and Knowledge Discovery*, *10*(3), e1355. https://doi.org/10.1002/widm.1355

Siantuba, J., Nkhata, L., & de Jong, T. (2023). The impact of an online inquiry-based learning environment addressing misconceptions on students' performance. *Smart Learning Environments*, *10*(1), 22. https://doi.org/10.1186/s40561-023-00236-y

Singh, J., & Singh, D. (2024). A comprehensive review of clustering techniques in artificial intelligence for knowledge discovery: Taxonomy, challenges, applications and future prospects. *Advanced Engineering Informatics*, *62*, 102799. https://doi.org/10.1016/j.aei.2024.102799

Tinajero, C., & Páramo, M. F. (1997). Field dependence-independence and academic achievement: a re-examination of their relationship. *British Journal of Educational Psychology*, *67*(2), 199–212. https://doi.org/10.1111/j.2044-8279.1997.tb01237.x

Wani, A. A. (2024). Comprehensive analysis of clustering algorithms: exploring limitations and innovative solutions. *PeerJ Computer Science*, *10*, e2286. https://doi.org/10.7717/peerj-cs.2286

Witkin, H. A., Moore, C. A., Goodenough, D., & Cox, P. W. (1977). Field-Dependent and Field-Independent Cognitive Styles and Their Educational Implications. *Review of Educational Research*, *47*(1), 1–64. https://doi.org/10.3102/00346543047001001

Zhang, J., Yang, Y., & Ding, J. (2023). Information criteria for model selection. *WIREs Computational Statistics*, *15*(5), e1607. https://doi.org/10.1002/wics.1607

Zhang, L., & Sternberg, R. J. (2005). A Threefold Model of Intellectual Styles. *Educational Psychology Review*, *17*(1), 1–53. https://doi.org/10.1007/s10648-005-1635-4