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. 3, July 2026, pp. 737-751

Structural Equation Modeling Based Partial Least Square of Student Misconceptions in Estimating Probability Distribution Parameters

Wardhani Utami Dewi[1*], Tri Cahyo Wahyudi[1], Miftahul Irfan[2]
[1]Universitas Muhammadiyah Metro, Indonesia
[3]Institut Teknologi Sumatera, Indonesia
wardhaniutamidewi@ummetro.ac.id

Article History:
Received : 17-02-2026
Revised : 15-03-2026
Accepted : 25-03-2026
Online : 01-07-2026

Keywords:
Misconception;
Parameter Estimation;
Probability Distribution;
SEM-PLS.

ABSTRACT

Students' misconceptions in estimating the parameters of probability distribution are still an important problem in learning statistics in universities. Most previous studies have examined partial misconceptions from cognitive aspects, so the structural relationship between concept understanding, learning experience, learning motivation, and problem-solving skills in explaining misconceptions has not been widely analyzed in an integrated manner. This study aims to develop and validate a structural model that explains the factors that influence student misconceptions in estimating probability distribution parameters using the Partial Least Squares Structural Equation Modeling (PLS-SEM) approach. This study uses an explanatory quantitative design involving 200 students who have studied probability distribution and parameter estimation in several universities in Lampung Province. Data was collected through a Likert scale questionnaire that measured five latent constructs, namely concept comprehension, learning experience, learning motivation, problem-solving skills, and student misconceptions. The analysis was carried out through the evaluation of the measurement model (loading factor, composite reliability, and average variance extracted), structural model testing using the bootstrapping technique, and evaluation of the overall suitability of the model. The results showed that concept comprehension ($\beta = 0.74$) and learning experience ($\beta = 0.82$) had a significant effect on problem-solving skills. Problem-solving skills further affect learning motivation ($\beta = 1.92$), while learning motivation affects the level of student misconception ($\beta = 0.67$). The developed model was able to explain 65% of the variation in problem-solving skills and 88% of the variation in student misconceptions. This research contributes in the form of a SEM-PLS model that integrates cognitive and affective factors in explaining the emergence of statistical misconceptions, as well as providing an empirical basis for the development of more effective statistical learning strategies to reduce student misconceptions.

https://doi.org/10.31764/jtam.v10i3.38398

———————— ◆ ————————

A. INTRODUCTION

Probability distributions, both discrete and continuous, are central to statistical education and play a fundamental role in parameter estimation. Mastery of probability distributions is essential not only for success in introductory probability courses in higher education but also for informed decision-making across various scientific and practical domains. For example, the Poisson distribution is used to predict the arrival of a patient in the emergency department (De Santis et al., 2022), the Binomial distribution is used to estimate vaccine efficacy (Senn, 2022), the Exponential distribution is used to model the lifespan of industrial components (Bilal et al., 2021), and the Normal distribution is widely applied in industrial quality control processes (Astuti et al., 2020; He & Xin, 2025; Khakifirooz et al., 2021).

Despite the important role of probability distributions in statistics, numerous studies indicate that students still experience significant difficulties in understanding probability distributions at a conceptual level. These difficulties often appear in the form of misconceptions when students are required to determine the appropriate distribution based on data characteristics, interpret the meaning of distribution parameters, or correctly apply parameter estimation procedures. Such misconceptions may lead to errors in statistical reasoning, including misinterpretations of variance (Shaw et al., 2025), confidence intervals (Sotos et al., 2007), and hypothesis testing results (Brus, 2021; Hubbard et al., 2019; Tong, 2019). Therefore, misconceptions in statistics not only affect students' conceptual understanding but may also influence the validity and reliability of scientific research findings.

Several studies have identified various factors contributing to the emergence of statistical misconceptions. Cognitive factors such as weak understanding of basic mathematical concepts and low statistical literacy are often reported as major causes of errors in understanding probability concepts (Korkmaz & Alkan, 2023; Yun et al., 2016). In addition, affective factors such as learning motivation, attitudes toward mathematics, and students' confidence levels also influence how students process and interpret statistical information (Soeharto & Csapó, 2021). Even students in STEM fields, who generally possess stronger analytical abilities, may still experience probabilistic reasoning errors such as base-rate neglect and the conjunction fallacy when dealing with complex probability problems (He & Xin, 2025).

Although various studies have addressed the factors that cause statistical misconceptions, most studies still examine these factors separately or focus only on one specific dimension, such as understanding concepts or attitudes towards learning mathematics. As a result, the structural relationship between cognitive factors, learning experiences, problem-solving skills, and affective factors in explaining the emergence of misconceptions has not been widely analyzed in an integrated manner. Therefore, an empirical model is still needed that is able to explain simultaneously how the interaction between these factors affects the emergence of student misconceptions in estimating probability distribution parameters. The development of structural models that integrate these various factors is important to gain a more comprehensive understanding of the mechanisms of statistical misconceptions and as a basis for designing more effective statistical learning strategies in higher education.

To address this research gap and analyze the complex relationships among multiple variables simultaneously, advanced multivariate analytical approaches are required. SEM is widely recognized as a powerful analytical method capable of examining complex relationships among latent variables. Various recent studies show the ability of SEM and PLS-SEM in modeling complex educational phenomena. For example, a study in Ghana found that the emotional intelligence (EI) of math teachers had a significant effect on teaching effectiveness, with gender as a moderator variable, although job satisfaction and self-efficacy did not play a mediator role (Owusu & Arthur, 2025). Another study in India used PLS-SEM to examine the entrepreneurial intentions of polytechnic students and found that peer influence, curriculum, and institutional support play an important role in shaping attitudes, subjective norms, and perceptions of behavioral control (Razi-ur-Rahim et al., 2024). In addition, studies using the Felder-Silverman model show that learning styles and anxiety levels also influence students' attitudes towards online learning, emphasizing the importance of adaptive and personalized educational design (Zagulova et al., 2023).

Cross-national studies also demonstrate contextual differences in the application of SEM in educational environments. Students in Malaysia benefit more from school support, while students in Indonesia show higher leadership and collaboration skills. This indicates that cultural support and learning activities have a different impact on the development of soft skills (Sobri et al., 2019). Then Riadi et al. (2024) used SEM to analyze the influence of motivation, self-confidence, and engagement on students' academic achievement in six public high schools in Denpasar, with results showing that motivation has a direct effect on achievement, while self-confidence and engagement affect achievement indirectly through motivation. Innovations in learning design have also been widely studied, for example, in Thailand, cloud-based learning models have been proven to significantly improve students' critical thinking skills and information literacy (Hongphanut, 2023).

SEM is an analytical approach that is able to model complex relationships between several latent variables simultaneously. SEM provides an analytical framework capable of modeling direct and indirect effects simultaneously, as well as accounting for measurement errors, thus providing a more comprehensive understanding of the complexity of misconceptions (Wei et al., 2025). However, conventional covariance-based SEM has limitations, including requiring large sample sizes and multivariate normality assumptions that are difficult to meet in education research (Gaskin et al., 2025). PLS-SEM is a more flexible variance-based approach that is suitable for small sample sizes, non-normal data, and complex models containing reflective and formative constructs (Demir & Uşak, 2025; Dewi et al., 2024; Zamir et al., 2022). Therefore, the PLS-SEM approach is suitable for modeling the relationship between concept understanding, learning experience, learning motivation, problem-solving skills, and student misconceptions.

Based on these considerations, this study aims to develop and validate a structural model that explains the factors influencing students' misconceptions in estimating probability distribution parameters using the PLS-SEM approach. The developed model integrates cognitive factors (understanding concepts and problem-solving skills), learning experience, and affective factors (learning motivation) to provide a more comprehensive understanding of the emergence of misconceptions in statistical learning. The main contribution of this research is the development of a structural model based on PLS-SEM that explains statistical misconceptions through the simultaneous interaction between cognitive and affective factors. The results of this study are expected to enrich the study of misconceptions in statistical learning and provide practical implications for the development of more effective learning strategies in reducing student misconceptions in higher education.

B. METHODS

This study uses a quantitative approach with the type of causal explanatory research because it aims to explain the cause-and-effect relationship between exogenous latent variables and endogenous latent variables (Bollen & Paxton, 1998) in the context of estimating probability distribution parameters. The population of this study is university students in Lampung Province, Indonesia. The research sample was selected using *the purposive sampling* technique, selected based on the criteria, namely, students in the second semester who have received probability distribution materials and parameter estimation. The indicators used were 14 indicators, with a sample of 200 students. According to Hair et al. (2021) , PLS-SEM can be applied using the 10-times rule, which suggests that the minimum sample size should be ten times the maximum number of structural paths directed at a construct. In this study, the most complex construct receives four structural paths, indicating a minimum sample requirement of 40 respondents. Therefore, the sample size used in this study exceeds the minimum requirement and is considered sufficient for SEM analysis. There are 2 endogenous latent variables, namely student misconceptions in estimating the distribution parameters probability (η_1) and problem-solving skills (η_2). Then, 3 exogenous latent variables, including the understanding of mathematical concepts (ξ_1), Learning experience (ξ_2), and learning motivation (ξ_3).

The questionnaire items were developed based on previous studies related to statistical misconceptions and learning factors in mathematics education. Each construct was operationalized into several measurable indicators reflecting theoretical dimensions of the variable. Prior to data collection, the instrument was reviewed by experts in statistics education to ensure content validity and clarity of the measurement items. This scale is used to assess the extent to which respondents agree or disagree with the statements submitted regarding the variables under study (Alabi & Jelili, 2023). All variables used *are fixed*, meaning that good indicator variables have been tested in systematically describing a construct (factor) using CFA. The following research variables are presented in Table 1.

Table 1. Research Variables

Laten Variables	Indicator Variables	
Concept Understanding (CU) / ξ_1	Understanding of basic mathematical concepts	X_1
	Understanding of the definition and concept of opportunity functions of discrete and continuous distributions	X_2
Learning Experience (LE) / ξ_2	Previous experience in studying statistics and probability topics	X_3
	The number of lecture hours that have been taken in statistics or probabilities	X_4
	Students' activeness in learning activities (e.g., discussions, assignments, and exams)	X_5
Motivation to learn (ML)/ ξ_3	Desire to learn more about statistics and probability	X_6
	Students' perception of the relevance of statistical material in their lives	X_7
	Positive attitude towards math learning	X_8
Problem Solving Skills (PSS)/ η_1	Ability to prove the postulates of probability distributions such as averages, variances, and momentary functions	Y_1
	Ability to solve statistical problems related to parameter estimation in daily life	Y_2

Laten Variables	Indicator Variables	
Student Misconceptions (SM)/ η_2	Errors in estimating parameters on discrete and continuous probability distributions	Y_3
	Inability to apply the parameter estimation formula used with the problems contained in the problem	Y_4
	Incomprehension in classifying between discrete and continuous probability distributions	Y_5

The conceptual framework of thinking can be designed and visualized in the form of a path diagram as shown in Figure 1.

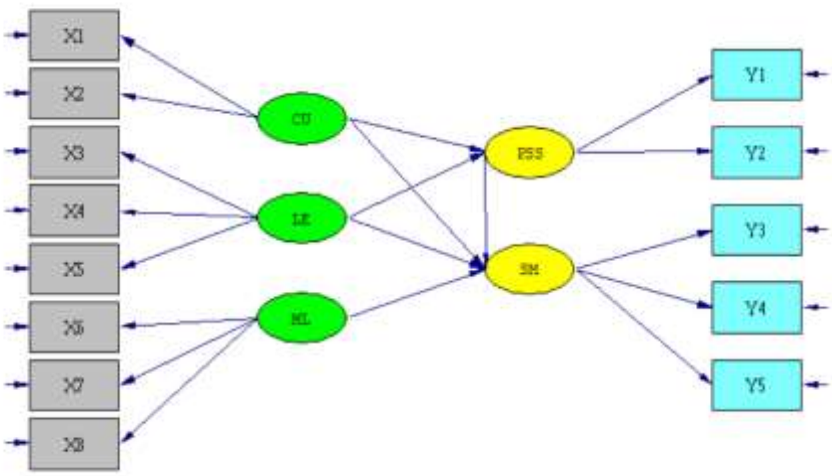


Figure 1. Conceptual Diagram (*SmartPLS software*)

Research Steps Conducted:

1. Model specifications, based on Figure 1. The specifications of the formed models are:
   a. Model structural

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1$$
$$\eta_2 = \beta_{21}\eta_1 + \gamma_{22}\xi_1 + \gamma_{22}\xi_2 + \gamma_{23}\xi_3 + \zeta_2$$

   b. Measurement Model

$$\begin{aligned} X_1 &= \lambda_{X_{11}}\xi_1 + \delta_1 & Y_1 &= \lambda_{Y_{11}}\eta_1 + \varepsilon_1 \\ X_2 &= \lambda_{X_{21}}\xi_1 + \delta_2 & Y_2 &= \lambda_{Y_{21}}\eta_1 + \varepsilon_2 \\ X_3 &= \lambda_{X_{31}}\xi_2 + \delta_3 & Y_3 &= \lambda_{Y_{31}}\eta_2 + \varepsilon_3 \\ X_4 &= \lambda_{X_{41}}\xi_2 + \delta_4 & Y_4 &= \lambda_{Y_{41}}\eta_2 + \varepsilon_4 \\ X_5 &= \lambda_{X_{51}}\xi_2 + \delta_5 & Y_5 &= \lambda_{Y_{51}}\eta_2 + \varepsilon_5 \\ X_6 &= \lambda_{X_{61}}\xi_3 + \delta_6 \\ X_7 &= \lambda_{X_{71}}\xi_3 + \delta_7 \\ X_8 &= \lambda_{X_{81}}\xi_3 + \delta_8 \end{aligned}$$

2. Construction of the Diagaram Line
   Constructing a path diagram is the process of building a representation of causal relationships between latent variables.
3. Parameter Estimation
   Parameter estimation was conducted using SmartPLS software, which is widely used for PLS-SEM The analysis includes measurement model evaluation and structural model testing using the bootstrapping procedure.
4. Evaluation of measurement models
   The evaluation of the measurement model was carried out by loading factor testing to test the validity of the variables. Correlation is said to satisfy the validity of the convergent if it has *a loading factor* or correlation coefficient $\lambda \geq 0{,}5$. Then perform CR (Composite Reliability) and AVE (Average Variance Extracted) tests to test the reliability of the variables. The latent variable can be said to have excellent reliability if the CR is more significant than 0.7.
5. Structural Model Evaluation
   The structural model was evaluated by examining path coefficients and their statistical significance obtained through bootstrapping. The explanatory power of the model was assessed using the coefficient of determination (R^2). A path relationship was considered significant when the t-value exceeded 1.96 at the 5% significance level.
6. Overall Model Evaluation
   Evaluation is carried out after the measurement model and the structural model are significant. The GFI, AGFI, RMR and PGFI values are a reference in evaluating the model as a whole.

C. RESULT AND DISCUSSION

Based on the results of a questionnaire involving 200 respondents from various study programs in Lampung, it can be seen that the respondents are spread across several scientific fields such as Mathematics, Statistics, Informatics, Management, Retail Management, Vocational Education, Industrial Engineering, and Elementary School Teacher Education. The distribution of respondents can be seen in Figure 2.

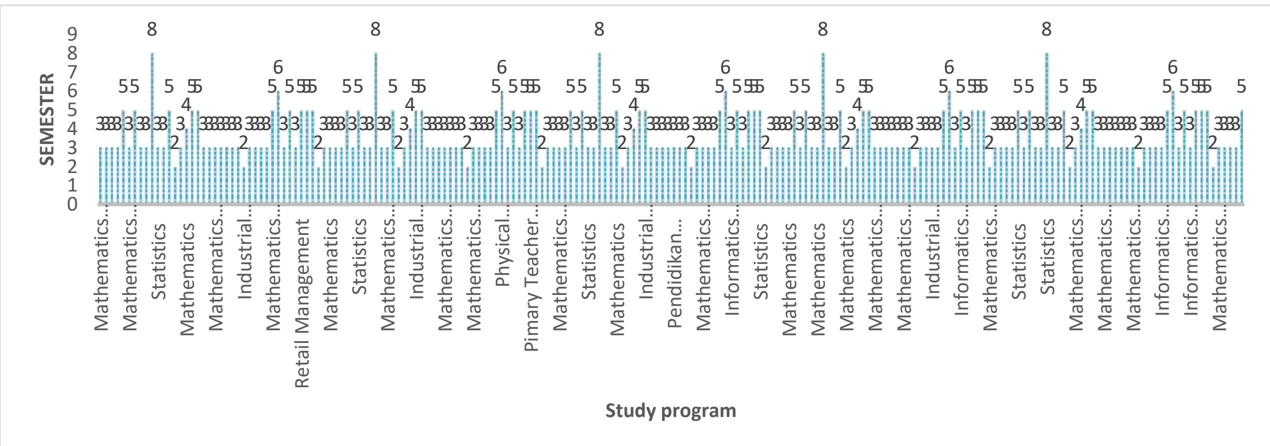
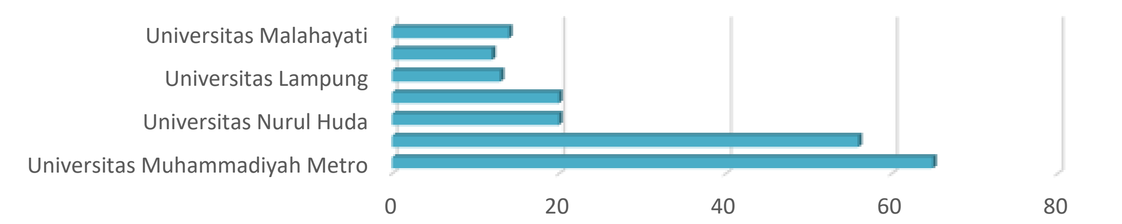


Figure 2. The distribution of respondents based on study programs and semesters that are effective

Based on the data distribution in Figure 2. It can be seen that the Mathematics study program accounts for the largest number of respondents compared to the other study programs. When viewed from the semester level, respondents came from semester 2 to semester 8, with a fairly strong dominance in semester 6 and semester 8. This shows that students who are in the middle to late stages of study are more involved in filling out questionnaires, possibly because the students have already taken Probability or Statistics courses, as well as similar courses that require data analysis skills. Meanwhile, students from non-science study programs such as Management, Retail Management, and Elementary School Teacher Education also participated, although the number was relatively smaller. In general, these results illustrate the fairly even representation of respondents between semesters and study programs, with a tendency to dominate from the science group, especially Mathematics and Statistics. Then, based on the number of distributions from the origin of Higher Education presented in Figure 3.



	Universitas Muhammadiyah Metro	Universitas Nahdlatul Ulama Lampung	Universitas Nurul Huda	Universitas Indonesia Mandiri	Universitas Lampung	Institut Teknologi Sumatera	Universitas Malahayati
Total	65	56	20	20	13	12	14

Figure 3. The distribution of respondents is based on the origin of universities in Lampung, Indonesia

Based on Figure 3. participants who participated in filling out the questionnaire were spread across seven Universities in Lampung. The highest number of respondents came from the Universitas Muhammadiyah Metro with a total of 65 people, followed by Universitas Nahdlatul Ulama Lampung with 56 people. These two institutions make the greatest contribution, so it can be said that it is quite representative of the student population in the Lampung area, especially from private universities based on community organizations.

1. Measurement Model

The measurement model test is carried out to ensure that the indicators used have validity and reliability in reflecting latent constructs. The results of the analysis show that all indicators have a fairly high loading factor, with most being above 0.70. This indicates that these indicators can adequately explain the constructed measure. The following is a path diagram of the measurement model in Figure 4 and Figure 5.

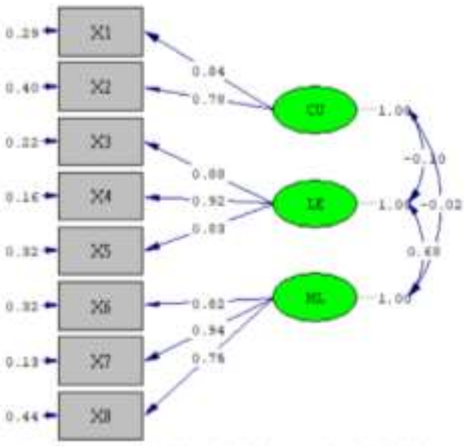


Figure 4. X-Model Measurement

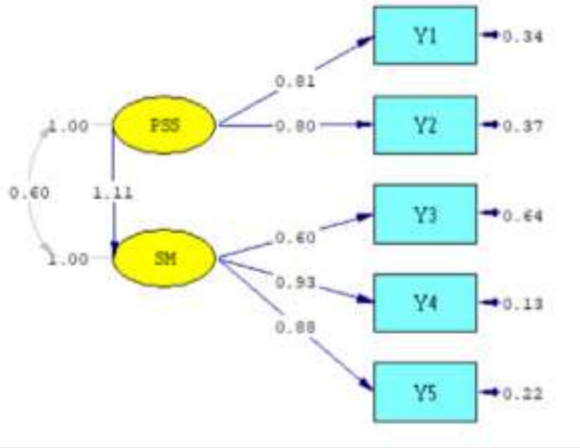


Figure 5. Y-Model Measurement

For Figure 4 the path diagram form of the measurement model with the indicator variable is variable X. While Figure 5 the path diagram of the measurement model with the indicator variable is variable Y. For more details on the loading factor value of each measurement model, it is shown in Table 2.

Table 2. Standardized regression weight (p-value = 0.000)

Relationship	Loading Factor	Standard Error (SE)	Composite Reliability (CR)	Average Variance Extracted (AVE)
$\xi_1 \rightarrow X_1$	0.84	0.29	0.79	0.65
$\xi_1 \rightarrow X_2$	0.78	0.40		
$\xi_2 \rightarrow X_3$	0.88	0.22	0.89	0.85
$\xi_2 \rightarrow X_4$	0.92	0.16		
$\xi_2 \rightarrow X_5$	0.83	0.32		
$\xi_3 \rightarrow X_6$	0.82	0.32	0.96	0.95
$\xi_3 \rightarrow X_7$	0.94	0.13		
$\xi_3 \rightarrow X_8$	0.75	0.44		
$\eta_1 \rightarrow Y_1$	0.81	0.34	0.78	0.64
$\eta_1 \rightarrow Y_2$	0.80	0.37		
$\eta_2 \rightarrow Y_3$	0.60	0.64	0.70	0.66
$\eta_2 \rightarrow Y_4$	0.93	0.13		
$\eta_2 \rightarrow Y_5$	0.88	0.22		

Note: p-value < 0.05

The results of the measurement model test in Table 2. show that the entire construct has a loading factor value that is above the minimum limit of 0.60, so that it can be declared valid convergently. In the ξ_1 construct reflected by the X1 and X2 indicators, it has a loading of 0.84 and 0.78, respectively, with an AVE value of 0.65 and a CR of 0.79. This shows that the ξ_1 construct is valid and reliable. The ξ_2 constructs measured by the X3, X4, and X5 indicators have very high loading (0.88, 0.92, and 0.83), with AVE 0.85 and CR 0.89, so this construct can be stated very well in terms of validity and reliability. Similarly, the ξ_3 construct reflected by X6, X7, and X8 exhibits strong loading (0.82, 0.94, and 0.75) with an AVE of 0.95 and a CR of 0.96, indicating a very high internal consistency.

On the endogenous construct side, the η_1 measured by Y1 and Y2 had loads of 0.81 and 0.80, respectively, with AVE of 0.64 and CR of 0.78, so it can be declared valid and reliable. Meanwhile, η_2 reflected by Y3, Y4, and Y5 shows a variation in the loading value, namely 0.60, 0,93, and 0.88, with AVE 0.66 and CR 0.70. Although the loading value on Y3 is relatively low, it still meets the minimum validity limit. Thus, overall, the η_2 construct can still be declared reliable even though it is at a minimum limit. Overall, these results show that all constructs in the model meet the criteria of convergent validity (AVE ≥ 0.50) and construct reliability (CR ≥ 0.70).

2. Structural Model

Structural model evaluation was carried out to identify and test the relationship between exogenous and endogenous latent variables. The following are the results of the structural model estimation presented in Figure 6.

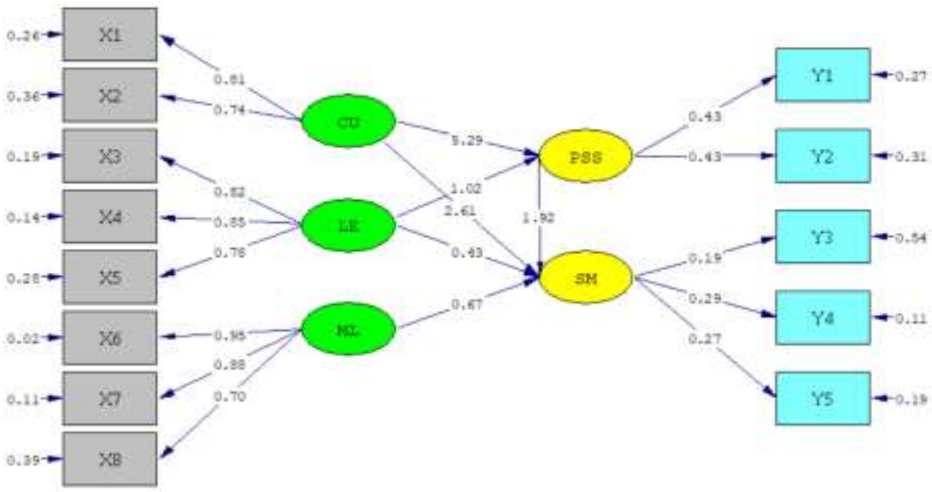


Figure 6. Estimation of Structural Model Parameters

Results from Figure 6 show how several factors are interconnected to affect students' motivation and skills. A good understanding of basic mathematical concepts has a significant effect on problem-solving skills (PSS), with a coefficient of 0.74. In addition, a good learning experience (LE) also has a great influence on this skill, with a coefficient of 0.82. Problem-solving skills alone have a huge impact on learning motivation, with a coefficient of 1.92. This means that the better the student's problem-solving skills, the higher their motivation to learn. A positive learning experience also contributes greatly to student motivation, with a coefficient of 0.85. Although student misconceptions affect several factors, the effect on learning motivation is smaller, with a coefficient of 0.67. Overall, this model shows that a deep understanding of basic concepts, positive learning experiences, and problem-solving skills are mutually supportive in increasing students' motivation and reducing misconceptions that can hinder their development.

Then, from the estimation results, a structural model test was carried out. Structural model evaluation was conducted by examining the significance of path coefficients using the bootstrapping procedure in SmartPLS. The significance of the relationships between variables was determined based on the t-statistics obtained from bootstrapping. A relationship is considered statistically significant when the t-statistic is greater than 1.96 at the significance level of $\alpha = 0.05$. In addition, the explanatory power of the structural model was evaluated using the coefficient of determination (R^2), which indicates a strong and reliable relationship between the variables in the structural model being tested. The following are the results of the evaluation of the structural model based on the t-values presented in Figure 7.

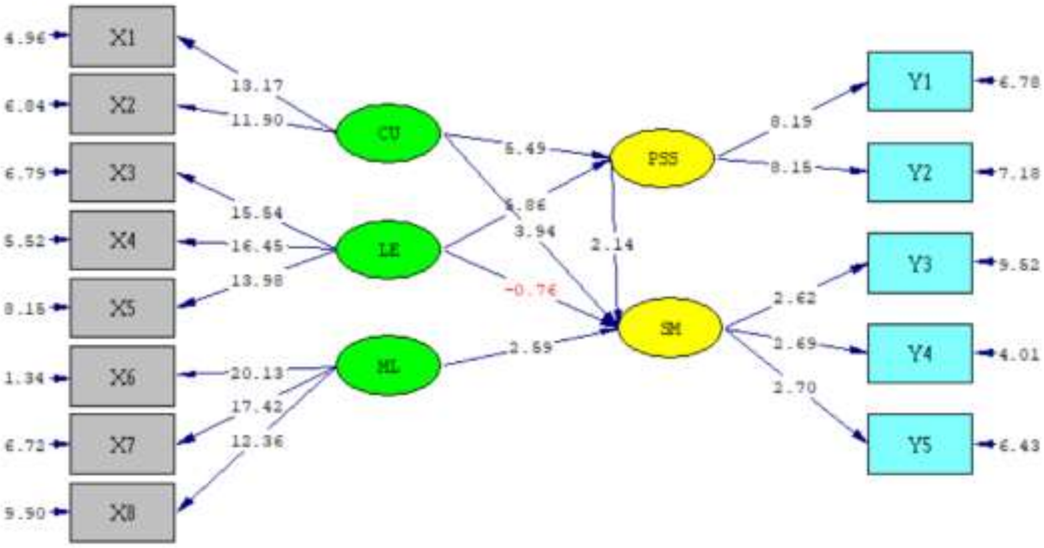


Figure 7. T-Values Diagram Path

Based on Figure 7, the t-statistics obtained from the bootstrapping procedure indicate the significance of the structural relationships between endogenous and exogenous latent variables. For more details, see Table 3.

Table 3. Structural Model Evaluation

Variable	Parameter Estimation (β)	Standard Errors	t-statistics	R-Squared
	$\eta_1 = 5.29\xi_1 + 1.02\xi_2$			0.65
ξ_1	5.29	1.34	3.94	
ξ_2	1.02	0.17	5.86	
	$\eta_2 = 1.92\,\eta_1 - 2.61\xi_1 - 0.43\xi_2 + 0.67\xi_3$			0.88
η_1	1.92	0.89	2.16	
ξ_1	-2.61	1.34	-1.95	
ξ_2	-0.43	0.57	-0.76	
ξ_3	0.67	0.26	2.59	

Based on Table 3. the first model, the relationship between CU and LE (Learning Environment) with PSS, obtained an estimated parameter of 5.29 for CU and 1.02 for LE. Both parameters are statistically significant, with t-statistics of 3.94 and 5.86, respectively, suggesting that both have a strong influence on PSS, which explains about 65% of the data variability. This means that the first model is quite good at describing the relationship between these factors and the PSS, although there is still about 35% of the variability that cannot be explained.

In the second model, which links PSS, CU, LE, and ML with SM, variable parameter estimates were found. PSS has a significant positive influence on SM with an estimated value of 1.92 and t-statistic of 2.16. On the other hand, CU and LE showed a negative effect on SM, but they were not statistically significant (t-statistic was -1.95 and -0.76, respectively), which means the effect could not be ascertained. In contrast, ML with an estimated parameter of 0.67 and t-statistic of 2.59 had a significant positive influence on SM. With an R^2 of 0.88, the second model can explain 88% of the variation in SM, making it particularly powerful in describing the factors that influence student motivation. Overall, although both models have important contributions, the second model is much more effective in explaining SM variability than the first model, which only explains PSS. This suggests that the ML factor becomes a very important determinant in student motivation, whereas CU and LE are more relevant in explaining PSS, although the influence is smaller in the second model.

3. Overall Model Evaluation

To ensure that the developed model can provide accurate and valid results, it is necessary to evaluate the model as a whole. This evaluation aims to assess the extent to which the model fits into the existing data, as well as identify areas that need improvement so that the model can better describe the relationships between variables. By conducting a comprehensive evaluation, we can ensure that the model meets the relevant goodness of fit criteria and can provide deeper and more precise insights in the context of the research or analysis being conducted. The following are the results of the model evaluation based on the LISREL output presented in Table 4.

Table 4. Overall Model Evaluation

Criterion	Value	Information
Root Mean Square Residual (RMR)	0.08	Excellent: An RMR value below 0.10 indicates a good model, the lower the better, and 0.08 indicates a model that fits the data.
Standardized RMR	0.13	Pretty Good: Although slightly higher than 0.10, it is still within acceptable limits and shows a slight inconsistency.
Goodness of Fit Index (GFI)	0.92	Excellent: Value GFI > 0.90 indicates that the model has an excellent fit with the data.
Adjusted Goodness of Fit Index (AGFI)	0.95	Excellent: AGFI Value > 0.90 shows that this model is excellent and has an excellent adjustment to the complexity of the model.
Parsimony Goodness of Fit Index (PGFI)	0.46	A PGFI value lower than 0.50 indicates that this model is more complex than necessary for a good fit.

Based on Table 4 the tested model showed an excellent match with the data. An RMR value of 0.08 indicates that the model has a fairly low average error, indicating that the model is relatively good at describing the data. Although the Standardized RMR value is slightly higher at 0.13, it is still within acceptable limits, which indicates a slight mismatch between the model and the data. However, the two main indicators, namely GFI (0.92) and AGFI (0.95), show excellent results, as they are both greater than 0.90, which means that the model is very good at matching data and considering the complexity of the model well. Although a PGFI of 0.46 indicates that the model may be a little more complex than necessary, this does not detract from the overall fit of the model. Thus, this model can be considered excellent, although there is little room for simplification.

4. Discussion

The results showed a clear cause-and-effect relationship between understanding basic mathematical concepts, learning experiences, problem-solving skills, and misconceptions experienced by students. A deeper understanding of basic mathematical concepts has a significant effect on students' problem-solving abilities, which in turn encourages increased motivation to learn. A positive learning experience, which involves active involvement in discussions and assignments, also helps students reduce misconceptions. This study provides a new view linking various cognitive and affective factors that influence student misconceptions, which have not been explored much in previous studies.

The main finding of this study is that students who have a better understanding of basic concepts tend to have better problem-solving skills as well. This has a positive impact on their motivation to learn. On the other hand, misconceptions in estimating probability distribution parameters were found to affect students' learning motivation, although the effect was smaller compared to other factors such as concept understanding and learning experience. This misconception generally arises from a misunderstanding of correct concepts and procedures, which hinders students from applying their statistical knowledge in practical situations.

Some of the factors that cause this result include the level of understanding of basic concepts possessed by students, learning experiences that involve active learning activities, and learning motivation that comes from within students. College students with a stronger understanding of basic concepts tend to have better problem-solving skills, which ultimately reduces misconceptions. Learning experiences that involve in-depth discussions and challenging tasks also have an important role in reducing misconceptions. In addition, students' learning motivation, which is greatly influenced by internal factors, also plays an important role in their learning outcomes.

The advantage of this study lies in the use of the SEM-PLS approach, which allows the analysis of complex relationships between variables simultaneously, as well as its ability to overcome measurement errors in the data. The study also integrates cognitive and affective factors, providing a more holistic view of student misconceptions. However, this study also has some shortcomings, such as the limitation of the sample that only includes students in Lampung Province, which may not fully represent the condition of students in other regions. In addition, the measurement of learning motivation in this study can be expanded to include more aspects that affect student learning outcomes.

When compared to previous studies, the results of this study are in line with the findings of Guerra-Reyes et al. (2024), which emphasize that a good understanding of basic concepts can prevent misconceptions in students. These findings are also similar to the results of the study (Casas del Rosal et al., 2024; Kurudirek et al., 2025), which states that misconceptions in statistics are largely due to a lack of understanding of concepts. However, this study offers a more comprehensive approach by combining cognitive and affective factors in the SEM-PLS model, which was less common in previous studies.

The implications of this research are very important for the development of higher education, especially in statistics and probability courses. These findings can be used to design more effective learning interventions, focusing not only on improving conceptual understanding but also on developing students' problem-solving skills and learning motivation.

The SEM-PLS model produced in this study can be a useful tool to analyze and overcome misconceptions in other fields of education, contribute to the development of educational research, and enrich the understanding of statistical misconceptions among students.

D. CONCLUSION AND SUGGESTIONS

The results of the structural model analysis indicate that students' conceptual understanding and problem-solving skills have a statistically significant role in reducing misconceptions in estimating probability distribution parameters, while positive learning experiences show a relatively weaker influence within the model. Learning motivation also contributes to reducing misconceptions, although its influence is slightly lower compared to conceptual understanding and problem-solving abilities. Using the SEM-PLS approach, this study successfully reveals the complex relationships among cognitive and affective factors that influence the occurrence of misconceptions in statistics learning. These findings provide important implications for statistics education in higher education, particularly in designing instructional strategies that emphasize conceptual understanding, analytical problem-solving activities, and meaningful learning experiences rather than focusing solely on procedural calculations. Therefore, the results of this study may serve as a reference for developing more holistic learning strategies that integrate cognitive, affective, and problem-solving aspects in order to reduce students' misconceptions in estimating probability distribution parameters.

Suggestions for future research include exploring the use of longitudinal designs to examine how misconceptions develop and change over time. A longitudinal approach would help determine whether improvements in conceptual understanding and problem-solving skills have a sustained effect on reducing misconceptions. Furthermore, methodological refinements are also recommended. Researchers could compare SEM-PLS with other advanced modeling approaches such as CB-SEM, multi-group analysis (MGA), or machine learning techniques.

ACKNOWLEDGEMENT

Thank you to the Institute for Research and Community Service (LPPM) of Universitas Muhammadiyah Metro for providing financial support under the Novice Lecturer Research Grant scheme No. 222/II.3.AU/C/LPPM/2025

REFERENCES

Alabi, A. T., & Jelili, M. O. (2023). Clarifying likert scale misconceptions for improved application in urban studies. *Quality & Quantity*, *57*(2), 1337–1350. https://doi.org/10.1007/s11135-022-01415-8

Astuti, D., Anggraeni, L., & Setyawan, F. (2020). Mathematical probability: student's misconception in higher education. *Journal of Physics: Conference Series*, *1613*(1), 12009. https://doi.org/10.1088/1742-6596/1613/1/012009

Bilal, M., Mohsin, M., & Aslam, M. (2021). Weibull-Exponential Distribution and Its Application in Monitoring Industrial Process. *Mathematical Problems in Engineering*, *2021*(1), 6650237. https://doi.org/10.1155/2021/6650237

Bollen, K. A., & Paxton, P. (1998). Interactions of latent variables in structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *5*(3), 267–293. https://doi.org/10.1080/10705519809540105

Brus, D. J. (2021). Statistical approaches for spatial sample survey: Persistent misconceptions and new developments. *European Journal of Soil Science*, *72*(2), 686–703. https://doi.org/10.1111/ejss.12988

Casas del Rosal, J. C., Muñoz Ñungo, O., Maz Machado, A., & Cuida Gómez, M. A. (2024). Approaching Statistics as a Distinct Subject and Its Impact on Reasoning and Conceptual Errors. *Revista de Gestão Social e Ambiental*, *18*(4), e06843. https://doi.org/10.24857/rgsa.v18n4-119

De Santis, A., Giovannelli, T., Lucidi, S., Messedaglia, M., & Roma, M. (2022). Determining the optimal piecewise constant approximation for the nonhomogeneous Poisson process rate of Emergency Department patient arrivals. *Flexible Services and Manufacturing Journal*, *34*(4), 979–1012. https://doi.org/10.1007/s10696-021-09408-9

Demir, S., & Uşak, M. (2025). Analyzing the Implementation of PLS-SEM in Educational Technology Research: A Review of the Past 10 Years. *SAGE Open*, *15*(2), 21582440251345950. https://doi.org/10.1177/21582440251345950

Dewi, W. U., Nisa, K., Setiawan, E., & Yudanegara, R. A. (2024). Structural equation modeling with generalized structured component analysis on the degree of public health in Indonesia. *AIP Conference Proceedings*, *2970*(1), 30006. https://doi.org/10.1063/5.0208230

Gaskin, J. E., Lowry, P. B., Rosengren, W., & Fife, P. T. (2025). Essential Validation Criteria for Rigorous Covariance-Based Structural Equation Modelling. *Information Systems Journal*, 6, 1630–1661. https://doi.org/10.1111/isj.12598

Guerra-Reyes, F., Guerra-Dávila, E., Naranjo-Toro, M., Basantes-Andrade, A., & Guevara-Betancourt, S. (2024). Misconceptions in the learning of natural sciences: A systematic review. *Education Sciences*, *14*(5), 497. https://doi.org/10.3390/educsci14050497

Hair, J. F., Astrachan, C. B., Moisescu, O. I., Radomir, L., Sarstedt, M., Vaithilingam, S., & Ringle, C. M. (2021). Executing and interpreting applications of PLS-SEM: Updates for family business researchers. *Journal of Family Business Strategy*, *12*(3), 100392. https://doi.org/10.1016/j.jfbs.2020.100392

He, S., & Xin, Y. P. (2025). Transforming fifth graders' misconceptions in solving probability comparison problems with purposefully designed instructional interventions. *Mathematical Thinking and Learning*, 1–23. https://doi.org/10.1080/10986065.2024.2403044

Hongphanut, R. (2023). A development of history instructional model on cloud technology to enhance critical thinking abilities and information literacy of undergraduate students. *Kasetsart Journal of Social Sciences*, *44*(3), 911–920. https://doi.org/10.34044/j.kjss.2023.44.3.29

Hubbard, R., Haig, B. D., & Parsa, R. A. (2019). The Limited Role of Formal Statistical Inference in Scientific Inference. *American Statistician*, *73*(sup1), 91–98. https://doi.org/10.1080/00031305.2018.1464947

Khakifirooz, M., Tercero-Gómez, V. G., & Woodall, W. H. (2021). The role of the normal distribution in statistical process monitoring. *Quality Engineering*, *33*(3), 497–510. https://doi.org/10.1080/08982112.2021.1909731

Korkmaz, Ebru, & Alkan, Selçuk. (2023). The Preservice Teachers' Approaches Toward Incorrect Probability Problems. *Sage Open*, *13*(4), 21582440231206604. https://doi.org/10.1177/21582440231206605

Kurudirek, A., Karim, B., Sarhang, D., & Tulqin, S. (2025). Math misconceptions: Mistakes, misunderstanding, and confusion. *Educenter: Jurnal Ilmiah Pendidikan*, *4*(1), 16–25. https://doi.org/10.55904/educenter.v4i1.1322

Owusu, D., & Arthur, F. (2025). Modelling the nexus between mathematics teachers' emotional intelligence and instructional effectiveness: a mediation moderation model using PLS-SEM approach. *Acta Psychologica*, *257*(June), 105119. https://doi.org/10.1016/j.actpsy.2025.105119

Razi-ur-Rahim, M., Uddin, F., Dwivedi, P., & Pandey, D. K. (2024). Entrepreneurial intentions among polytechnic students in India: Examining the theory of planned behaviour using PLS-SEM. *The International Journal of Management Education*, *22*(3), 101020. https://doi.org/10.1016/j.ijme.2024.101020

Riadi, A., Sukarsa, I., Suciptawati, N., & Kencana, E. (2024). Causalities Of Motivation, Self-Related Beliefs, And Student Engagement Towards High School Students' Achievement Using Structural Equation Modeling. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *18*(4 SE-Articles). https://doi.org/10.30598/barekengvol18iss4pp2307-2318

Senn, S. (2022). The design and analysis of vaccine trials for COVID-19 for the purpose of estimating efficacy. *Pharmaceutical Statistics*, *21*(4), 790–807. https://doi.org/10.1002/pst.2226

Shaw, E., Lin, J. M., & Natarajan, P. (2025). Navigating Statistical Inference: Addressing Enduring Misconceptions in Social Science Hypothesis Testing. *International Journal of Social Sciences, Language and Linguistics*, *5*(04), 6–9. https://doi.org/10.55640/ijssll-05-04-02

Sobri, K. M., Hanum, F., Zulnaidi, H., & Ahmad, A. R. (2019). A comparative study of school environment for students' skills development in Malaysia and Indonesia. *Kasetsart Journal of Social Sciences*, *40*(1), 149–154. https://so04.tci-thaijo.org/index.php/kjss/article/view/239829

Soeharto, S., & Csapó, B. (2021). Evaluating item difficulty patterns for assessing student misconceptions in science across physics, chemistry, and biology concepts. *Heliyon*, *7*(11), e08352. https://doi.org/https://doi.org/10.1016/j.heliyon.2021.e08352

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review*, *2*(2), 98–113. https://doi.org/10.1016/j.edurev.2007.04.001

Tong, C. (2019). Statistical Inference Enables Bad Science; Statistical Thinking Enables Good Science. *American Statistician*, *73*(sup1), 246–261. https://doi.org/10.1080/00031305.2018.1518264

Wei, G., Wang, G., Wijaya, T. T., Su, M., & Chen, J. (2025). Exploring the factors influencing student problem posing in a mathematics classroom: PLS-SEM and fuzzy set qualitative comparative analysis. *Acta Psychologica*, *259*, 105339. https://doi.org/https://doi.org/10.1016/j.actpsy.2025.105339

Yun, H. J., Ko, E.-S., & Yoo, Y. J. (2016). *Students' Misconceptions and Mistakes Related to Measurement in Statistical Investigation and Graphical Representation of Data BT - The Teaching and Learning of Statistics: International Perspectives* (D. Ben-Zvi & K. Makar (eds.); pp. 119–120). Springer International Publishing. https://doi.org/10.1007/978-3-319-23470-0_14

Zagulova, D., Prokofyeva, N., Yulia, K., Ziborova, V., Katalnikova, S., & Popova, Y. (2023). Personalization of Distance Learning: Using PLS-SEM for Analyzing the Possibility of Differentiating Students with the Felder-Silverman Learning Style Model. *Procedia Computer Science*, *225*, 3077–3085. https://doi.org/https://doi.org/10.1016/j.procs.2023.10.301

Zamir, S., Yang, Z., Wenwu, H., & Sarwar, U. (2022). Assessing the attitude and problem-based learning in mathematics through PLS-SEM modeling. *PLoS ONE*, *17*(5 May), 1–15. https://doi.org/10.1371/journal.pone.0266363