

Unraveling Chemometric Analysis Performance: A SEM-PLS Investigation of Simple Technology Implementation, Statistical Understanding, and Outlier Detection

**Muhammad Rizky¹, Fitri Aditri¹, Ibrahim Fanji Dipura¹, Regina Aulia Ramadhani¹
Hulwatul Adzro¹, Neni Alyani^{1,2}, M Miftahul Madya¹**

¹Lembaga Riset AI Creation (LRAC), Depok, Indonesia

²Institut Pemerintahan Dalam Negeri, Sumedang, Jawa Barat

*Email corresponding author: mmiftahulm29@gmail.com

Received 7 March 2026

Accepted 1 April 2026

Available Online 30 April 2026

ABSTRACT: This study aims to examine the effect of Simple Technology Implementation on spectral preprocessing accuracy using the Savitzky–Golay method, with data literacy as a moderating variable. A quantitative approach was employed using the Structural Equation Modeling–Partial Least Squares (SEM-PLS) method. Data were collected from 113 respondents using a five-point Likert scale questionnaire. The results indicate that Simple Technology Implementation has a positive effect on improving descriptive statistical understanding, although its impact on outlier detection effectiveness remains limited. The measurement model reveals that several constructs do not fully meet validity and reliability criteria, particularly the data literacy variable. These findings suggest that while user-friendly technology can enhance analytical efficiency, it is not sufficient to support more complex analytical tasks. Therefore, improving analytical performance requires both accessible technology and stronger user competencies in data literacy and statistical understanding.

ABSTRAK: Penelitian ini bertujuan untuk menganalisis pengaruh Simple Technology Implementation terhadap akurasi preprocessing spektra menggunakan metode Savitzky–Golay dengan literasi data sebagai variabel moderasi. Pendekatan kuantitatif digunakan dengan metode Structural Equation Modeling–Partial Least Squares (SEM-PLS). Data dikumpulkan dari 113 responden menggunakan kuesioner skala Likert lima poin. Hasil penelitian menunjukkan bahwa implementasi teknologi sederhana berpengaruh positif terhadap peningkatan pemahaman statistika deskriptif, namun pengaruhnya terhadap efektivitas deteksi outlier masih terbatas. Model pengukuran menunjukkan bahwa beberapa konstruk belum sepenuhnya memenuhi kriteria validitas dan reliabilitas, khususnya variabel literasi data. Temuan ini mengindikasikan bahwa teknologi yang mudah digunakan mampu meningkatkan efisiensi analisis, tetapi belum cukup untuk mendukung tugas analisis yang lebih kompleks. Oleh karena itu, peningkatan kinerja analisis memerlukan teknologi yang aksesibel serta penguatan kompetensi literasi data dan pemahaman statistik pengguna.

Keywords: Chemometrics; Data Literacy; SEM-PLS; Simple Technology Implementation; Spectral Preprocessing

1. INTRODUCTION

The development of visual-based analytical technologies has driven a transformation in data analysis processes, particularly in the field of chemometrics. Platforms such as Orange Data Mining enable users, including novice research communities, to perform data analysis more easily through graphical user interfaces and drag-and-drop workflow systems. The implementation of simple technology in such

platforms not only reduces technical complexity but also accelerates data exploration and interpretation processes. Therefore, ease of use becomes a crucial factor in improving the effectiveness of chemometric-based data analysis (Demšar et al., 2013).

In the context of chemometric analysis, the spectral preprocessing stage, particularly using the Savitzky–Golay filter method, plays a critical role in enhancing data quality prior to further analysis. Proper

preprocessing can reduce noise, clarify spectral patterns, and improve the accuracy of data interpretation. However, the success of preprocessing is not solely determined by the method employed but also by the user's ability to understand the data, including the ability to detect outliers and comprehend descriptive statistics such as mean and data distribution. Thus, the implementation of simple technology is expected to support improved preprocessing quality by enhancing users' data understanding (Savitzky & Golay, 1964).

Several previous studies have examined the impact of technology on data analysis using the Structural Equation Modeling–Partial Least Squares (SEM-PLS) approach. SEM-PLS is widely used due to its capability to analyze complex relationships among latent variables, including direct, mediating, and moderating effects (Hair et al., 2017). Empirical studies demonstrate that SEM-PLS is effective in evaluating technology adoption, user behavior, and analytical performance across various domains, including education, digital technology utilization, and data-driven decision-making. For instance, research in educational contexts highlights how digital technology acceptance and sustainability are influenced by perceived ease of use, effectiveness, and user interest (Pratama et al., 2025), while studies in chemometric and analytical domains reveal that perception and knowledge significantly contribute to the development of analytical competence (Alyani et al., 2025a). This is further supported by recent research indicating that conceptual understanding, including knowledge of validation objectives, validation parameters, and chemometric principles, significantly influences the implementation of analytical methods, highlighting the importance of cognitive variables in analytical practices (Nursyafaah et al., 2026a). Furthermore, SEM-PLS has also been widely applied to examine complex social and organizational phenomena, such as leadership, cultural intelligence, and social adaptation (Alyani et al., 2025b), demonstrating its robustness in modeling multidimensional constructs. In the context of work readiness and technology utilization, previous studies have shown that

factors such as program quality, mentor support, and the use of digital tools contribute significantly to improving performance outcomes (Adzro et al., 2026). Other findings also highlight that these relationships are strengthened through the presence of moderating and mediating variables (Dipura et al., 2026). Furthermore, additional evidence suggests that these supporting elements play an essential role in enhancing overall work readiness (Nursyafaah et al., 2026a). In a similar vein, research in the field of digital business demonstrates that technological readiness is a key determinant of technology adoption behavior (Nursyafaah et al., 2026b). Other studies emphasize that digital capability contributes significantly to improving organizational performance (Alyani et al., 2026a). Moreover, perceived ease of use has been identified as an important factor influencing the adoption of artificial intelligence technologies (Alyani et al., 2026b). Additionally, broader theoretical perspectives confirm that these factors collectively shape user acceptance and system utilization (Venkatesh et al., 2003). In addition, recent research emphasizes the importance of user awareness, service quality, and motivation in shaping behavioral outcomes and system utilization, further reinforcing the flexibility of SEM-PLS in integrating behavioral and technological constructs (Alyani et al., 2026c). These findings collectively suggest that the SEM-PLS approach is highly suitable for examining the relationships among technology implementation, statistical understanding, and data analysis effectiveness within complex analytical environments.

Despite the growing body of literature on the application of SEM-PLS in technology adoption and data analysis, several research gaps remain. First, prior studies tend to examine variables such as technology implementation, user understanding, and analytical performance in isolation, without integrating them into a unified analytical framework, particularly in the context of chemometric analysis. Second, although previous research has highlighted the importance of statistical understanding and

data interpretation skills, limited attention has been given to how these factors simultaneously contribute to improving analytical outcomes, especially in terms of preprocessing accuracy. Third, while data literacy has been acknowledged as an important factor in technology utilization, its role as a moderating variable in strengthening the relationship between simple technology implementation and analytical performance remains underexplored.

Therefore, this study aims to fill these gaps by developing an integrated model that examines the effect of *Simple Technology Implementation* on spectral preprocessing accuracy, represented by outlier detection effectiveness and descriptive statistical understanding, with community data literacy as a moderating variable. This study is expected to provide both theoretical contributions by extending the application of SEM-PLS in chemometric contexts and practical implications by offering insights into improving data analysis performance through accessible and user-friendly technology.

2. METHODOLOGY

This research adopts a quantitative approach utilizing the Structural Equation Modeling–Partial Least Squares (SEM-PLS) technique to investigate the relationships among the proposed variables. The SEM-PLS method is chosen due to its flexibility in handling complex research models and its capability to simultaneously analyze multiple relationships, including moderating effects (Hair et al., 2011). Moreover, this method is particularly suitable for studies with relatively limited sample sizes and does not impose strict assumptions regarding data normality, making it appropriate for exploratory and predictive research contexts (Chin, 1998).

Data collection was conducted through a survey administered to 113 respondents who are familiar with or have experience using Orange Data Mining. The instrument employed a five-point Likert scale, ranging from strongly disagree (1) to strongly agree (5), to capture participants'

perceptions of the constructs under investigation. The study involves three main variables: *Simple Technology Implementation* as the independent variable, *Spectral Preprocessing Accuracy (Savitzky–Golay)* as the dependent variable, and *Data Literacy* as the moderating variable. Detailed indicators for each construct are organized and presented in tabular form to enhance clarity and avoid redundancy in the narrative description.

Table 1 Variable, Definition, and Indicator

Variable	Type	Code	Indicator
X	Independent	X1	I can install Orange Data Mining easily
		X2	The visual interface of Orange is easy for new users to understand
		X3	The process of inputting analytical data into Orange is easy to perform.
		X4	I can create an analysis workflow using the drag-and-drop system easily.
		X5	The chemometric analysis features in Orange are easy to use.
		X6	The analysis results displayed by Orange are easy to understand.
Y1	Dependent	Y1.1	The smoothing method helps to refine the spectrum.
		Y1.2	Noise in the spectrum is reduced after preprocessing.
		Y1.3	Spectral peaks become clearer after preprocessing.
		Y1.4	The results of spectral data preprocessing are sufficiently stable.
Y2	Dependent	Y2.1	Orange helps me understand the mean of the data.
		Y2.2	I can understand data variability through the displayed standard deviation.

		Y2.3	Data distribution visualization makes data interpretation easier.
		Y2.4	Statistical graphs help me understand the characteristics of the data.
Z	Moderator	Z1	I Understand the basic concepts of data analysis.
		Z2	I am able to read graphs and data visualizations effectively.
		Z3	I am capable of interpreting data analysis results.
		Z4	I have experience using data analysis software.

The analysis procedure follows the standard SEM-PLS framework, beginning with the evaluation of the measurement model (outer model). This stage focuses on assessing the validity and reliability of the constructs. Convergent validity is examined through outer loading values and Average Variance Extracted (AVE), while internal consistency reliability is evaluated using Cronbach's Alpha and composite reliability (rho_c) (Kock, 2015). Additionally, the Variance Inflation Factor (VIF) is utilized to detect potential multicollinearity among indicators, ensuring that the measurement model meets the required statistical criteria (Henseler et al., 2015).

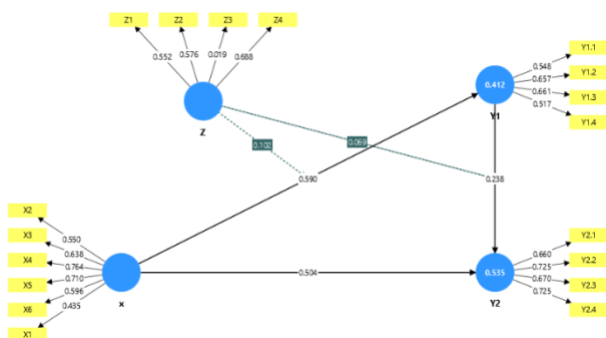


Fig. 1 Model Diagram and Intersubject Relationship

Subsequently, the structural model (inner model) is assessed to determine the relationships between variables and to test the proposed hypotheses. Particular

attention is given to the moderating role of data literacy in influencing the relationship between simple technology implementation and spectral preprocessing accuracy. The significance of the hypothesized paths is evaluated using bootstrapping techniques, which generate statistical measures such as path coefficients, t-values, and p-values. This analytical approach ensures that the findings are statistically robust and capable of explaining the structural relationships within the proposed research model.

3. RESULT

The measurement model evaluation was conducted to assess the validity and reliability of the constructs using outer loading, Average Variance Extracted (AVE), Cronbach's Alpha, composite reliability (rho_c), and Variance Inflation Factor (VIF). The results show that the construct of *Simple Technology Implementation (X)* does not fully meet the criteria for convergent validity, as indicated by an AVE value of 0.390, which is below the recommended threshold of 0.50. In terms of reliability, the construct shows a Cronbach's Alpha value of 0.678 and composite reliability (rho_c) of 0.789, indicating marginal reliability (Hair et al., 2017).

At the indicator level, X4 (0.764) and X5 (0.710) meet the validity threshold and are considered valid indicators. Meanwhile, X2 (0.550), X3 (0.638), and X6 (0.596) show moderate loading values that are still acceptable. However, X1 (0.435) is below the minimum threshold, indicating that it is not valid and may reduce the overall quality of the construct. All VIF values for the X construct are below 5, indicating no multicollinearity issues (Kock, 2015).

For the dependent variable, the construct is represented by two dimensions, Y1 and Y2. The Y1 dimension shows weak measurement results, with an AVE value of 0.359 and low reliability (Cronbach's Alpha = 0.398; rho_c = 0.689), indicating that it does not meet validity and reliability criteria. In contrast, the Y2 dimension shows better results, with an AVE value of 0.484 (close to the threshold) and reliability values (Cronbach's Alpha = 0.644; rho_c = 0.789) indicating moderate

internal consistency. All indicators in Y2 have loading values above 0.60, suggesting acceptable validity.

The moderating variable, *Data Literacy (Z)*, shows weak measurement results, with AVE = 0.278, Cronbach's Alpha = 0.340, and rho_c = 0.538. These values indicate that the construct is neither valid nor reliable. Indicator Z3 has an extremely low loading value (0.019), making it clearly invalid, while Z1, Z2, and Z4 show moderate but insufficient loadings. However, all VIF values remain below the critical threshold, indicating no multicollinearity issues in the model.

Table 2 Result outer models

Construct & Indicator	Loading Factor	AVE	Cronbach's Alpha	rho_c	VIF
X		0.390	0.678	0.789	
X1	0.435				1.119
X2	0.550				1.135
X3	0.638				1.321
X4	0.764				1.611
X5	0.710				1.512
X6	0.596				1.207
Y1		0.359	0.398	0.689	
Y1.2	0.548				1.183
Y1.2	0.657				1.175
Y1.3	0.661				1.121
Y1.4	0.517				1.090
Y2		0.484	0.644	0.789	
Y2.1	0.660				1.247
Y2.2	0.725				1.374
Y2.3	0.670				1.185
Y2.4	0.725				1.323
Z		0.278	0.340	0.538	
Z1	0.552				1.108
Z2	0.576				1.138
Z3	0.019				1.058
Z4	0.688				1.065
Z x Y1	1.000				1.000
Z x X	1.000				1.000

Construct & Indicator	Description
X	Not valid, not reliable
X1	Not valid
X2	Valid
X3	Valid
X4	Valid
X5	Valid
X6	Valid
Y1	Valid & reliable

Y1.1	Valid
Y1.2	Valid
Y1.3	Valid
Y1.4	Valid
Y2	Valid & reliable
Y2.1	Valid
Y2.2	Valid
Y2.3	Valid
Y2.4	Valid
Z	Not valid & not reliable
Z1	Valid
Z2	Valid
Z3	Not valid
Z4	Valid
Z x X	Valid

4. DISCUSSION

The findings indicate that the implementation of simple technology contributes positively to data analysis processes, particularly in enhancing users' understanding of statistical outputs. This result supports prior studies emphasizing that ease of use and intuitive system design are critical factors in improving user performance and facilitating technology adoption (Venkatesh et al., 2003). In the context of Orange Data Mining, the visual interface and drag-and-drop workflow enable users to process and interpret data more efficiently, especially in understanding descriptive statistical results.

However, the results also show that the influence of technology implementation is not uniform across all analytical dimensions. The weak measurement performance of the Y1 dimension suggests that tasks such as outlier detection are more complex and cannot be fully supported by system simplicity alone. This indicates that certain analytical processes require deeper cognitive skills and domain knowledge beyond what is provided by user-friendly tools. In other words, while technology can enhance accessibility, it does not entirely replace analytical expertise.

Furthermore, the moderating role of data literacy was not effectively demonstrated in this study due to the weak validity and reliability of the construct. Although data literacy is conceptually expected to strengthen the relationship between technology implementation and analytical performance, the empirical results suggest

that its measurement needs further refinement. The low loading values indicate that the current indicators may not adequately capture users' actual analytical capabilities.

Overall, these findings highlight that the effectiveness of simple technology implementation in chemometric analysis depends on the interaction between technological usability and user competence. While platforms such as Orange Data Mining can significantly simplify analytical workflows, the quality of the analysis remains influenced by users' understanding of statistical concepts and their ability to interpret data. Therefore, improving analytical performance requires not only accessible technology but also efforts to enhance data literacy and analytical skills among users.

5. CONCLUSION

This study demonstrates that *Simple Technology Implementation* has a positive role in supporting chemometric data analysis, particularly in improving users' understanding of descriptive statistical results. The findings indicate that user-friendly platforms such as Orange Data Mining can enhance analytical efficiency by simplifying workflows and data interpretation processes. However, the results also reveal that technology alone is not sufficient to ensure optimal analytical performance, especially for more complex tasks such as outlier detection. In addition, the moderating role of data literacy was not effectively captured due to measurement limitations, indicating the need for more robust indicators in future research. Therefore, improving chemometric analysis performance requires not only accessible technology but also strengthening users' data literacy and analytical competencies.

REFERENCES

- Adzro, H., et al. (2026). Analysis of interns' work readiness: The roles of program quality, mentor support, and project management tools using SEM-PLS. *J-GADS: Journal of Governance and Applied Data Science*, 1(1).
- Alyani, N., et al. (2025a). Analysis of perception and knowledge on chemometric competence among students and practitioners using SEM-PLS. *Jurnal Manajemen Sumber Daya Aparatur*, 13(2).
- Alyani, N., et al. (2025b). Building a culture of tolerance through sustainable leadership, cultural intelligence, and social adaptation. *Khazanah Sosial*, 7(4).
- Alyani, N., et al. (2026a). Service awareness and management effects on library visit frequency mediated by user interest using SEM-PLS. *J-GADS: Journal of Governance and Applied Data Science*, 1(1).
- Alyani, N., et al. (2026b). Pengaruh digital capability terhadap innovation capability dan operational efficiency startup melalui pemanfaatan AI generatif. *J-KBDI: Jurnal Kecerdasan Buatan dan Data Science Industri*, 1(1).
- Alyani, N., et al. (2026c). Pengaruh perceived ease of use terhadap AI generative adoption dalam meningkatkan competitive advantage startup digital. *J-KBDI: Jurnal Kecerdasan Buatan dan Data Science Industri*, 1(1).
- Chin, W. W. (1998). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern methods for business research* (pp. 295–336). Lawrence Erlbaum Associates.
- Demšar, J., et al. (2013). Orange: Data mining toolbox in Python. *Journal of Machine Learning Research*, 14, 2349–2353.
- Dipura, I. F., et al. (2026). The influence of promotion and social influence on online purchase decisions mediated by convenience. *J-GADS: Journal of Governance and Applied Data Science*, 1(1).
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2017). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Sage.

- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice*, 19(2), 139–152.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.
- Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration*, 11(4), 1–10.
- Nursyafaah, T. H., et al. (2026a). The effect of work environment and motivation on work readiness through competence using a SEM-PLS approach. *J-GADS: Journal of Governance and Applied Data Science*, 1(1).
- Nursyafaah, T. H., et al. (2026b). Pengaruh technology readiness terhadap adopsi AI generatif dan dampaknya terhadap kinerja startup menggunakan SEM-PLS. *J-KBDI: Jurnal Kecerdasan Buatan dan Data Science Industri*, 1(1).
- Nursyafaah, T. H., et al. (2026). Pemanfaatan SEM-PLS dalam menganalisis variabel kognitif terhadap validasi metode pada bidang kimia. *J-TP MIPA: Jurnal Transformasi Pendidikan MIPA*, 1(1).
- Pratama, M. O., et al. (2025). Analisis kebutuhan dan minat dalam pemanfaatan teknologi digital di SMK Depok menggunakan SEM-PLS. *Jurnal Terapan Informatika Nusantara*, 6(6), 635–643.
- Savitzky, A., & Golay, M. J. E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36(8), 1627–1639.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425–478.