CHAPTER 3
RESEARCH METHOD

This chapter outlines the research method which includes the research design, population and sample, research instruments, instrument testing, data collection procedures, data analysis.

A. Research Design

Research design is a systematic plan for investigating a scientific issue. It is intended to offer a suitable framework for research.[85] With the aims of investigating the influence of the Merdeka Curriculum on teacher creativity, professional development, and teacher burnout among EFL Teachers in this study employs a quantitative approach for data analysis and interpretation. The chosen research design for this study is a survey research design and analyze using Structural Equation Modeling (SEM).

Survey research designs are methods in quantitative research where researchers conduct surveys to a sample or the entire population to understand attitudes, opinions, behaviors, or characteristics. They gather numerical data through questionnaires or interviews and analyze it statistically to identify trends and test hypotheses. Researchers also interpret the data by comparing it to past studies.[86] Survey designs come in two types: cross-sectional and longitudinal. Since data for this study was gathered only once, a cross-sectional survey design was used. Cross-sectional survey designs are helpful for gathering data about "current attitudes, opinions, or beliefs," according to Creswell.[87]

This study employs SEM (Structural Equation Modeling) in analyzing the data. It is a statistical technique used to test causal relationships between latent variables developed based on theory. SEM combines factor analysis and regression or path analysis to test and confirm the relationship model between variables that

---

[85] Pankajakshan Vijayanthi Indu and K. Vidhukumar, "Research Designs- an Overview," *Kerala Journal of Psychiatry* 32, no. 1 (2019): 64, https://doi.org/10.30834/KJP.32.1.2019.179.
[86] John W. Creswell, *Educational Research: Planning, Conducting, and Evaluating Quantitative and Qualitative Research* (Pearson, 2012), 376.
[87] Creswell, 377.

have been predetermined.[88] Schumacker and Lomax [89] defined SEM as a technique used to test theories about how different variables are related. Researcher use SEM to see if their ideas about relationships between variables are supported by data.

B. Population and Sample

A population is a group of elements or cases, including individuals, objects, events, or occurrences, that have specific criteria and for which the research results are intended (Fraenkel J. R, and Wallen, N.E., as cited in Rasyid).[90] Creswell identified a target or sampling frame, as a group of individuals or organizations sharing a common characteristic that the researcher can identify and study.[91]

In this study, the target population is all English as a Foreign Language (EFL) teachers across Indonesia. These include teachers from various educational levels—elementary, junior high, senior high school, and vocational schools—both in public and private institutions. The teachers are actively involved in English language instruction and are expected to have experienced or currently be experiencing the implementation of the Merdeka Curriculum. The population also includes teachers who are members of professional teacher forums such as MGMP (Musyawarah Guru Mata Pelajaran), which tend to be more exposed to current curriculum updates and pedagogical training. By focusing on this group, the study aims to capture a comprehensive view of how curriculum changes contribute to teacher creativity, professional development, and potential burnout across diverse educational contexts in Indonesia.

Within this target population, the researcher then selects a sample for study. According to Creswell[92], a sample is a subgroup of the target population that the researcher plans to study for generalizing about the target population. In ideal situation, the sample consists of individuals who are representative of the entire

---

[88] Singgih Santoso, *Analisis Structural Equation Modelling (SEM) menggunakan AMOS 26* (Elex Media Komputindo, 2021), 1.

[89] Randall E. Schumacker and Richard G. Lomax, *A Beginner's Guide to Structural Equation Modeling* (Routledge, 2010), 2.

[90] Fathor Rasyid, *Metodologi Penelitian Kualitatif Dan Kuantitatif Teori, Metode Dan Praktek* (LPPM IAIN KEDIRI, 2022), 126, https://lppm.iainkediri.ac.id/metodologi-penelitian-kualitatif-dan-kuantitatif-teori-metode-dan-praktek.

[91] Creswell, *Educational Research, 142*.

[92] Creswell, 142.

population. Convenience Sampling is used in this study. Convenience sampling is a type of non-probability sampling where the nearest individuals are chosen as respondents, continuing this process until the required sample size is reached.[93] In convenience sampling, the researcher selects a sample from those who are easy to reach. This method was selected due to practical considerations such as time constraints, geographic spread, and online accessibility.

The final sample consists of EFL teachers who voluntarily responded to an online questionnaire distributed through teacher forums such as MGMP WhatsApp groups. While the sample is not randomly selected, efforts were made to ensure variation in respondents' backgrounds, such as teaching experience, school type, and geographical location, to reflect the broader characteristics of the population. All participants have been involved in the implementation of the Merdeka Curriculum to varying degrees, providing relevant insights for the aims of this study.

Moreover, According Ghozali the recommended sample size for SEM models requires a minimum sample of 100–200 respondents to ensure a more stable model estimation.[94] Since SEM uses the Maximum Likelihood method, which is covariance-based, it requires data to be normally distributed or at least approximately normal. A common ratio used is that for each parameter to be tested, there should be at least 15 data (samples or respondents).[95] Therefore, in this study, since there are 13 parameters, the minimum required sample size is 195 data.

C. Types and Source of the Data

The type of data used in this research is primary quantitative data. In research methodology, data are generally classified into two types: primary data and secondary data. According to Creswell[96], primary data are data collected directly by the researcher for the specific purpose of addressing the research

[93] Hossein Tavakoli, *A Dictionary of Research Methodology and Statistics in Applied Linguistics* (Rahnama Press, 2012).

[94] Imam Ghozali, *Model Persamaan Struktural, Konsep Dan Aplikasi Dengan Program AMOS 24*, 7th ed. (Badan Penerbit Universitas Diponegoro, 2017), 62, https://imamghozali.com/produk-29-.html.

[95] Santoso, *Analisis Structural Equation Modelling (SEM) menggunakan AMOS 26*.

[96] Creswell, *Educational Research*.

problem, while secondary data refer to data previously collected by others for different purposes. In this study, the data were obtained through a structured questionnaire developed to measure three latent variables: teacher creativity, professional development, and teacher burnout, in relation to the implementation of the Merdeka Curriculum.

The source of data is English as a Foreign Language (EFL) teachers across various regions in Indonesia, who are actively engaged in the teaching and learning process and have experienced the impact of curriculum changes. By collecting data directly from EFL teachers, the study ensures that the findings are grounded in real classroom contexts and reflect authentic teacher experiences.

D. Operational Definition of Variables

To understand Structural Equation Modeling, it is essential to comprehend the fundamental terms used in this type of research. In the basic concept of research, the term "variable" is frequently encountered. A variable is an object determined by the researcher to be studied, observed, and analyzed for conclusions.[97]

1. Latent Variables

Latent variables (constructs or factors) are variables that are not directly observable or measured. Instead, they are deduced from a set of observed variables that measured through tests, surveys, and similar methods.[98] There are two types of latent variables: exogenous (independent) and endogenous (dependence).

In a research model that involves several latent variables, there is also a need for a set of indicators (manifest), and there will be various relationships among the variables. Such a complex model is referred to as a SEM model.[99]

In this study, there are four latent variables; Merdeka curriculum as exogenous latent variable (ξ); teacher creativity, professional development and

---

[97] Zulkifli Musannip Efendi Siregar et al., *Structural Equation Modeling, Konsep Dan Implementasinya Pada Kajian Ilmu Manajemen Dengan Menggunakan Amos*, 1st ed. (DeePublish, 2021), 2.
[98] Schumacker and Lomax, *A Beginner's Guide to Structural Equation Modeling, 3*.
[99] Siregar et al., *Structural Equation Modeling, Konsep Dan Implementasinya Pada Kajian Ilmu Manajemen Dengan Menggunakan Amos, 2*.

teacher burnout as endogenous latent variables (η). To be observed it needs to break down those latent variables into manifest variables or observed variables. The breakdown for the manifest variables for each of them will be analyzed in the following sections.

1.1. Exogenous Latent Variables (Independent)

In this study Merdeka Curriculum serves as the exogenous latent variable. Latent variables cannot be observed directly and, therefore, cannot be measured directly. Consequently, a researcher needs to operationalize these latent variables to represent them accurately.[100] So, the manifest variables or indicators of Merdeka Curriculum are flexible learning (X_1); Essential Materials (X_2); Diverse Teaching (X_3); Technology Integration (X_4).

1.2. Endogenous Latent Variables (Dependences)

There are three endogenous latent variables for this study: teacher creativity, professional development and teacher burnout. These will be breaking down as follows:

a. Teacher Creativity serves as the first endogenous latent variable. The manifest variables of teacher creativity are Fluency (Y_1); Flexibility (Y_2); Elaboration (Y_3).

b. Professional Development represents the second endogenous latent variable in the model. The manifest variables or indicators of professional development are Teacher Professional Education (PPG) (Y_4), Training (Y_5) and *Guru Penggerak* (Y_6).

c. Teacher Burnout acts as the last endogenous latent variable that will be observed. The burnout's indicators are Workloads (Y_7); Administrative Tasks (Y_8) and Student Behavior (Y_9).

---

[100] Siregar et al, 3.

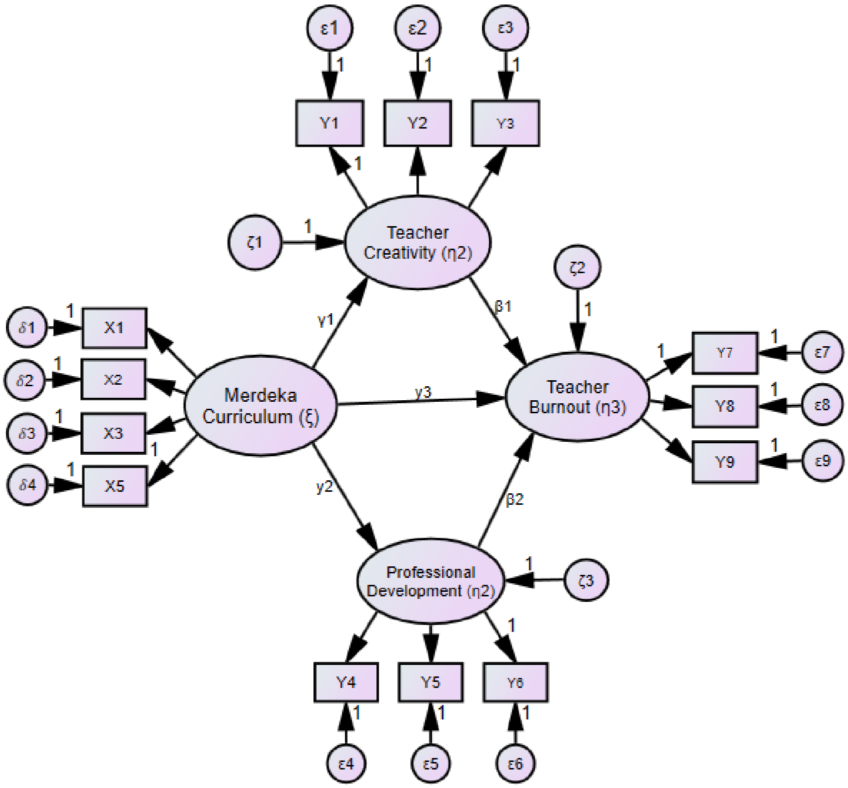

Figure 3. 1. Hypothesized model specification.

2. ***Observed Variables***

Manifest variables are variables that can be measured or directly observed (observed variables), or in other words, they serve as indicators of latent variables. For this study, there are three endogenous latent variables that each of them has their own observed variables that already broke down in the previous sections. The manifest variables used to form exogenous latent constructs are symbolized by ξ, while the manifest variables used to form endogenous latent constructs are symbolized by η.[101] Since there are three endogenous variables in this study, different symbols are employed for each manifest variable associated with each latent variable. For teacher creativity uses (η_1), Professional Development (η_2) and Teacher Burnout (η_3).

[101] Siswoyo Haryono, *METODE SEM Untuk Penelitian Manajemen AMOS, LISREL, PLS*, 1st ed. (Jakarta: PT. Intermedia Personalia Utama, 2016), 37.

E. Research Instruments

Research instruments are tools designed to gather various types of information in research. These tools facilitate the systematic collection, quantitative or qualitative processing, and organized arrangement of data.[102] Instrumentation for this study is questionnaire.

A questionnaire serves as an important tool in a research study to help the researcher in the collection of the relevant data related to the research topic.[103] Questions are like the language researchers use to find out what they need to know for their study. The answers people give to these questions help researchers find the information they are looking for. The questionnaire in this study was developed through a structured and theory-based process. It included the following steps: Identifying → Planning → Developing → Piloting (Pre-test) → Implementing.[104] This step-by-step approach was used to make sure the questionnaire was relevant to the research topic, easy to understand, and reliable for measuring the intended variables.

1. Identifying

The first stage involved identifying the core latent constructs to be measured, namely: curriculum changes, teacher creativity, professional development, and teacher burnout. Each construct was theoretically grounded. The dimension of curriculum changes was derived from the Merdeka Curriculum guidelines issued by the Directorate of Junior High Schools, Ministry of Education and Culture[105], Flexible Curriculum Structure, Focus on Essential Materials, Use Diverse Teaching Tools, Utilization of Digital Technology. Teacher creativity was operationalized using indicators from Guilford's[106] creativity theory, which

[102] Sudarmin Sukmawati, "Development of Quality Instruments and Data Collection Techniques," *Jurnal Pendidikan Dan Pengajaran Guru Sekolah Dasar (JPPGuseda)* 6, no. 1 (March 31, 2023): 119, https://doi.org/10.55215/jppguseda.v6i1.7527.

[103] Hamed Taherdoost, "Designing a Questionnaire for a Research Paper: A Comprehensive Guide to Design and Develop an Effective Questionnaire," SSRN Scholarly Paper (Rochester, NY: Social Science Research Network, August 1, 2022), https://papers.ssrn.com/abstract=4178678.

[104] Aek Phakiti, "Questionnaire Development and Analysis," in *The Companion to Language Assessment* (John Wiley & Sons, Ltd, 2013), 1245–61, https://doi.org/10.1002/9781118411360.wbcla068.

[105] "Kurikulum Merdeka."

[106] Guilford, *The Nature of Human Intelligence*.

includes originality, flexibility, fluency, and elaboration. The construct of professional development was informed by the national framework outlined in SIM PKB, highlighting Teacher Professional Education (PPG), Training Program and Guru Penggerak. Meanwhile, teacher burnout was conceptualized based on empirical studies conducted by Aeria et al. and Hastings and Bham, which emphasize to administrative tasks, workloads and students' behavior.

2. Planning

Following the theoretical definition of constructs, a detailed blueprint was designed. The blueprint mapped each theoretical dimension to several questionnaire items to ensure comprehensive coverage and content validity. Below is the breakdown of each variable along with its corresponding dimensions and item indicators:

Table 3. 1. Instrument blueprint for Merdeka Curriculum

Variable	Sub-Variables	Indicators	Items
Merdeka Curriculum (Dit-SMP Kemdikbud, 2024)	a. Flexible Curriculum Structure	Understand the flexible curriculum.	1
		Perceived impact of the flexible curriculum on student learning.	2
	b. Focus on Essential Materials	Clarity on essential materials in the curriculum.	3
		Relevance and usefulness of essential materials.	4
	c. Use of Diverse Teaching Tools	Frequency of incorporating diverse teaching tools.	5
		Comfort level with using various teaching tools.	9
	d. Utilization of Digital Technology	Frequency of digital technology use in the classroom.	6
		Proficiency with digital tools and technology.	7
		Impact of digital technology on student learning.	8

Table 3. 2. Instrument blueprint for Teacher Creativity

Variable	Sub-Variables	Indicators	Items
Teacher Creativity (Guilford, 1967)	a. Fluency	Come up with many ideas and solutions	1,5
	b. Flexibility	The way of thinking in facing situations or challenges	2
		Look at things from different angles, approaches, strategies.	3
	c. Elaboration	Make creative outputs interestingly and originally	4

Table 3. 3. Instrument blueprint for Professional Development

Variable	Sub-Variables	Indicators	Items
Professional Development (SIM PKB, 2024)	a. Teacher Professional Education (PPG)	Perception of PPG programs contribution	1
		Potential influence of PPG on student performance and teaching quality	2,3
	b. Training Program	Perception on training program contribution	4
		The potential influence of Training in teaching skills	5
	c. *Guru Penggerak* Program	Perception on the *Guru Penggerak* program contribution	6
		Teaching method of *Guru Penggerak* Program exploration	7
		The implementation of new teaching methods by *Guru Penggerak* participants.	8

Table 3. 4. Instrument blueprint for Teacher Burnout

Variable	Sub-Variables	Indicators	Items
Aeria, et al (2018); Hastings and Bam (2003)	a. Workloads	Number of teaching hours per week.	1
		Frequency of extra duties beyond teaching (e.g., extracurricular activities, meetings).	2
		Perceived sufficiency of time to prepare lessons.	3
		Perception of workload manageability.	4
	b. Administrative Tasks	Time spent on paperwork and administrative tasks.	5
		Frequency of non-teaching responsibilities (e.g., grading, reporting).	6

		Perceived impact of administrative tasks on teaching effectiveness.	7
	c. Student Behavior	Frequency of dealing with challenging student behavior.	8
		Perceived impact of student behavior on teaching and stress levels.	9, 10

3. Developing

Questionnaire items were developed based on theoretical indicators drawn from the literature. For instance, teacher creativity items reflected practical expressions of originality and flexibility in teaching approaches, while items on professional development addressed the extent of engagement in training or self-initiated learning. Curriculum change items focused on perceived shifts in pedagogical practices, and burnout items were designed to capture emotional strain and disengagement. All items were initially written in English and subsequently translated into Bahasa Indonesia, followed by back-translation to ensure semantic consistency. The finalized questionnaire resulting from the development process is presented in Appendix 1.

3.1. Validity Testing

According to Cronbach and Meehl[107], validity refers to the extent to which a test measures what it is intended to measure. A questionnaire is considered valid when its questions effectively capture the construct it aims to evaluate. In this study, expert validity was employed, utilizing Aiken's V technique for the validity test of the questionnaire instrument. This technique assesses various aspects, including content, construct, and language.[108]

[107] Lee J Cronbach and Paul E Meehl, "Construct Validity in Psychological Tests" 52 (1955): 281–302.

[108] Lewis R. Aiken, "Three Coefficients for Analyzing the Reliability and Validity of Ratings," *Educational and Psychological Measurement* 45, no. 1 (March 1, 1985): 131–42, https://doi.org/10.1177/0013164485451012.

Validators assign scores to each item based on a predetermined scale, and the obtained scores are subsequently analyzed using Aiken's V,[109] as follow:

Validity Each Item $$v = \frac{\sum s}{n\,(c-1)}$$

Overall Validity $$v = \frac{\sum s}{mn\,(c-1)}$$

s	= $r - l_0$
V	= index of expert agreement on the validity of the item
r	= expert choice category score
l_0	= lowest score in the scoring category
n	= number of experts
m	= number of items
c	= number of categories that experts can choose from

To determine if the instrument is categorized as valid, the validation coefficient is based on the following criteria[110]:

Table 3. 5. Aiken Validity Criteria

Criteria	Description
$V \leq 0.8$	High
$0.4 \leq V \leq 0.8$	Medium
$V < 0.4$	Low

The Aiken's V validation test was conducted based on the adaptation of the assessment forms from Sukarman[111] and Nabil, et al.[112], focusing on three key aspects: *Instructions, Language, and Content Suitability*.

Using the Aiken's V coefficient, an analysis was carried out to assess the validity of the questionnaire that had been developed. The results of this validity test are presented in the following table:

[109] Heri Retnawati, *Analisis Kuantitatif Instrumen Penelitian (Panduan Peneliti, Mahasiswa, dan Psikometrian)* (Parama Publishing, 2016), 18.
[110] Retnawati, 38.
[111] Sukarman Sukarman, "Analisis Pengaruh Motivasi Belajar Terhadap Hasil Belajar Fisika Peserta Didik dengan Menggunakan Metode SEM (Structural Equation Modeling) pada Kelas XI Matematika Ilmu Alam (MIA) SMA Negeri se-Kecamatan Bolo Kabupaten Bima" (diploma, Universitas Islam Negeri Alauddin Makassar, 2018), https://repositori.uin-alauddin.ac.id/11818/.
[112] Naimina Restu An Nabil et al., "Analisis Indeks Aiken untuk Mengetahui Validitas Isi Instrumen Asesmen Kompetensi Minimum Berbasis Konteks Sains Kimia," *PAEDAGOGIA* 25, no. 2 (September 2, 2022): 187, https://doi.org/10.20961/paedagogia.v25i2.64566.

Table 3. 6. Aiken's V Validity Test Results

No	Aspects	Rater 1	Rater 2	S_1	S_2	$\sum s$	V	Description	Categories
1.	Instruction	8	8	6	6	12	0.875	Valid	High
2.	Language	12	11	9	8	17	0.875	Valid	High
3.	Content	16	15	12	11	23	0.90	Valid	High
Overall		36	34	27	25	52	**0.72**	Valid	Medium

Source: Data Results 2025

Table 3.6 shows that the validation coefficient for the Instruction aspect is 0.875, for Language is 0.875, and for Content Feasibility is 0.90. This indicates that the instrument used is categorized as valid. The Aiken's V test allows for a comprehensive evaluation of the instrument's overall validity. The overall validity coefficient is 0.72, which falls within the criteria of $0.4 \leq V \leq 0.8$, categorizing it as medium validity. This indicates that the instrument is valid and appropriate for use in the study.

3.2.Reliability Testing

According to Imam Ghozali, reliability testing is a tool used to assess a questionnaire consisting of indicators for a variable or construct. In this study, which utilized expert validity, reliability testing was conducted using *Inter-Rater Reliability* (IRR). Inter-rater reliability measures the reliability based on the level of agreement between raters (evaluators) and provides a score that reflects the extent of consensus or agreement among the experts.

The result of the Cohen's Kappa (K) agreement coefficient, calculated using SPSS 29, is presented in the table below:

Symmetric Measures

		Value	Asymptotic Standard Error[a]	Approximate T[b]	Approximate Significance
Measure of Agreement	Kappa	.571	.229	2.619	.009
N of Valid Cases		9			

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

Figure 3. 2. The result of Cohen Kappa Agreement

According to Landis and Koch[113], the kappa coefficient values are categorized as follows:

Table 3. 7. Kappa Coefficient Values

Criteria	Description
$\kappa < 0.00$	Poor Agreement
$0.00 < \kappa < 0.20$	Slight
$0.21 < \kappa < 0.40$	Fair
$0.41 < \kappa < 0.60$	Moderate
$0.61 < \kappa < 0.80$	Substantial
$0.81 < \kappa < 1.00$	Almost perfect agreement

Based on the analysis results, the Cohen's Kappa coefficient is 0.571. According to Landis and Koch, a Kappa value of 0.571 falls within the moderate agreement category (0.41 – 0.60). With a Kappa value of 0.571, the reliability of the instrument falls within the moderate agreement range, meaning that the ratings given by the two raters are fairly consistent but may still have some differences.

4. Pilot Testing

Pilot study was carried out involving a group of 40 EFL teachers who shared characteristics with the study's target population but were excluded from the main data collection. The purpose of the pilot was to assess the clarity, readability, and practical applicability of the questionnaire, and to statistically evaluate its internal consistency. Feedback from respondents was used to confirm whether the items were comprehensible and appropriate within the respondents' teaching context. In addition, statistical analyses such as construct validity were used to determine the consistency and coherence of items within each construct.

1.1. Validity Testing

In this study, construct validity was tested using Factor Analysis with the assistance of SPSS 29. The results of the Kaiser-Meyer-Olkin (KMO) test and Bartlett's Test of Sphericity indicated that the data were suitable for factor analysis.

[113] Jacob Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement* 20, no. 1 (April 1960): 37–46, https://doi.org/10.1177/001316446002000104.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.631
Bartlett's Test of Sphericity	Approx. Chi-Square	1523.011
	df	465
	Sig.	<.001

Figure 3. 3. The result of KMO and Bartlett's Test

The KMO value was 0.631, which falls into the “mediocre” category, suggesting an acceptable level of sampling adequacy. Meanwhile, Bartlett’s Test was statistically significant ($\chi^2 = 1523.011$, df = 465, $p < .001$), indicating that the correlations among items were sufficient to proceed with factor extraction.

Based on the results of the Rotated Component Matrix (See Appendix), the factor loadings were generally well distributed across seven components. Most items showed loading values above 0.50, indicating that they contributed meaningfully to their respective factors. Some items, such as TB1 (0.911), TB2 (0.890), TB9 (0.908), and TB10 (0.888), had very high loadings, reflecting a strong and consistent factor structure for the Teacher Burnout construct. Likewise, several items from the Curriculum Change construct—such as MC5 (0.748), MC6 (0.746), MC7 (0.762), and MC8 (0.777)—also showed strong associations with their components.

However, a few items had relatively low loadings, such as MC4 (0.448) and TC5 (0.445), which fell below the commonly accepted threshold of 0.50. These items may not adequately represent their intended constructs. However, since both items were considered important in representing key aspects of their respective constructs, the researcher decided to retain them in the final instrument. This decision was made based on theoretical relevance, despite their loading values being slightly below the conventional threshold.

1.2.Reliability Testing

Reliability testing is a tool used to measure a questionnaire that consists of indicators representing a construct or variable. An item is considered reliable if its Cronbach’s Alpha value is greater than 0.60. The following table presents the results of the reliability test for the questionnaire used in this study.

Table 3. 8. The Result of Reliability Test

Variables	Cronbach Alfa	Description
Merdeka Curriculum Changes	.927	Reliable
Teacher Creativity	.878	Reliable
Professional Development	.913	Reliable
Teacher Burnout	.787	Reliable

Source: SPSS Primary Research Data 2025

The reliability test results, measured using Cronbach's Alpha, show that all four variables in this study meet the reliability criteria. The Merdeka Curriculum Changes variable achieved a Cronbach's Alpha value of 0.927, indicating an excellent level of internal consistency. The Teacher Creativity variable also demonstrated strong reliability, with a value of 0.878. Similarly, the Professional Development variable showed high reliability with a Cronbach's Alpha of 0.913. Although slightly lower, the Teacher Burnout variable also met the acceptable reliability threshold, with a Cronbach's Alpha of 0.787, which still indicates a good level of consistency. These results suggest that all variables in the questionnaire are statistically reliable and suitable for further analysis.

F. Data Collection

Data for this study was collected using self-administered questionnaires, where respondents filled out the forms independently without the presence of a researcher. This approach allowed participants to complete the questionnaire at their own convenience, encouraging more honest and reflective responses.[114][115]

The process of disseminating the questionnaire started with the researcher identifying and contacting the MGMP of English Teacher networks across Indonesia. The researcher made efforts to reach out to as many MGMP groups as possible, ensuring a broad representation from various regions.

---

114 Alan Bryman, *Social Research Methods*, Fifth edition (Oxford: Oxford University Press, 2016), https://ktpu.kpi.ua/wp-content/uploads/2014/02/social-research-methods-alan-bryman.pdf.
115 Don A. Dillman, Jolene D. Smyth, and Leah Melani Christian, "Internet, Phone, Mail, and Mixed-Mode Surveys," 4th edition (John Wiley & Sons, Inc, 2014), https://content.e-bookshelf.de/media/reading/L-2753682-49f7ffb446.pdf.

The questionnaire was then shared online using Microsoft forms (https://forms.office.com/r/L5EQrEhWi0) through the MGMPs WhatsApp group spread across the country. The dissemination was conducted digitally to facilitate easy access and participation from teachers in different locations. It aimed to collect diverse responses, especially from English teachers in different regions of Indonesia.

For the MGMP of Kediri District, they were given the opportunity to disseminate the questionnaire in person through online meeting, ensuring that local teachers could easily participate and provide their responses. This dissemination took place on February 4, 2025, through Zoom meeting during the regular MGMP meeting to share good practices. After all dissemination process, a total of 200 completed questionnaires were collected.

G. Data Processing

The data analysis in this study was done using IBM SPSS AMOS 26, which is part of the IBM SPSS software package. Although there are many programs available for Structural Equation Modeling (SEM), the researcher chose AMOS because it is beginner-friendly, easy to use, and does not require complex coding. It also works well with SPSS.

Before using SPSS or AMOS, the data first went through data cleaning and transcription. Out of 200 responses, 5 were removed because of duplicate entries, unclear personal information, or because the respondent did not meet the research criteria — for example, one respondent was only a school volunteer.

After cleaning, the next step was data transcription using Excel. In this step, verbal answers were converted into numbers so they could be analyzed. Since each variable had several questions, and AMOS only needs one number per variable, the average score of each variable was calculated.

Once the transcription and averaging were complete, the data was entered into IBM SPSS Statistics 29 to make sure the input was accurate and well-organized. After that, the data was moved into AMOS for the main SEM analysis.

H. Data Analysis Technique

1. Descriptive Statistical Analysis

Descriptive analysis is a method used to describe or summarize the object being studied based on sample or population data, as it appears, without intending to make broad conclusions.[116] In this study, the observed variables are X and Y, along with latent variables like the Merdeka Curriculum, Teacher Creativity, Professional Development, and Teacher Burnout. Descriptive analysis is also used to summarize the scores for all observed variables in this research. This process includes calculating the average score, determining the standard deviation, and categorizing the data.

To present respondents' perceptions of each variable, in this study, descriptive statistics are used with categories of Low, Moderate, and High. These categories are based on the mean scores and the theoretical range. The theoretical range is calculated using the formula:

$$\text{Interval} = \frac{Maximum\ Score - Minimum\ Score}{Number\ of\ Intervals}$$

The use of Low, Moderate, and High categories helps to simplify interpretation and allows for a clearer comparison of how strongly each variable is perceived or experienced by the respondents.

2. Inferential Statistical Analysis

Inferential analysis was conducted to draw conclusions about the population based on the data collected from the sample.[117] This type of analysis helps in making predictions, testing hypotheses, and understanding relationships between variables beyond the observed data. In this study, several inferential analyses were conducted to draw conclusions about the population based on the

[116] Nasir Nasir and Sukmawati Sukmawati, "Analysis of Research Data Quantitative and Qualitative," *Edumaspul: Jurnal Pendidikan* 7, no. 1 (March 1, 2023): 369.
[117] Nasir and Sukmawati, 369.

sample data. This included assessing normality, testing for validity and reliability. and hypothesis testing.

a) Assessment of Normality

Normality test was conducted to assess whether the data followed a normal distribution. This is an important step because many statistical analyses, including SEM, assume that the data is normally distributed. In SEM, data is typically considered normally distributed if the skewness values fall within ±2 and the kurtosis values do not exceed ±7, as suggested by Byrne[118] and West, Finch, and Curran[119].

b) Assessment of Validity

One of the key benefits of SEM is its ability to assess the construct validity of the proposed measurement model. Construct validity measures how well the indicator variables reflect the theoretical latent construct. In other words, construct validity provides confidence that the indicator scores taken from the sample accurately represent the true scores in the population. To assess validity, the Average Variance Extracted (AVE) is used. To evaluate convergent validity, one common measure used is Average Variance Extracted (AVE). AVE calculates how much of the variance in the indicators is explained by the construct itself. The formula for AVE is:

$$\text{Average Variance Extracted} = \frac{\sum \boldsymbol{\lambda i}^2}{\sum \boldsymbol{\lambda i}^2 + \sum \boldsymbol{\varepsilon}}$$

The symbol $\boldsymbol{\lambda}$ represents the standardized factor loading, and *i* refers to the number of items or indicators. For n items, AVE is calculated by dividing the sum of the squared standardized factor loadings (squared multiple correlations) by the sum of the squared standardized loadings plus the total error variance. An AVE value of 0.50 or higher indicates good convergent validity.

[118] Barbara M. Byrne, *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming, Second Edition*, 2nd ed. (New York: Routledge, 2013), https://doi.org/10.4324/9780203805534.

[119] Stephen G. West, John F. Finch, and Patrick J. Curran, "Structural Equation Models with Nonnormal Variables: Problems and Remedies," in *Structural Equation Modeling: Concepts, Issues, and Applications* (Thousand Oaks, CA, US: Sage Publications, Inc, 1995), 56–75.

c) Assessment of Reliability

Reliability is also one of the indicators used to assess convergent validity. While many researchers use Cronbach's Alpha as a measure of reliability, it is known to underestimate the true reliability compared to Construct Reliability (CR). To calculate reliability, Construct Reliability (CR) is used with the following formula:

$$\text{Construct Reliability (CR)} = \frac{(\sum \boldsymbol{\lambda i})^2}{(\sum \boldsymbol{\lambda i})^2 + \sum \boldsymbol{\varepsilon}}$$

A CR value of 0.70 or higher indicates good reliability, while values between 0.60 and 0.70 are still acceptable, provided that the indicators in the model demonstrate good validity.

3. Structural Equation Modelling (SEM)

In this study, the researcher will examine how the independent variables affect the dependent variable using the Structural Equation Modeling (SEM) method. The test conducted in the research measures the actual input fit with the prediction of the proposed model and hypothesis testing. In this study, IBM SPSS AMOS 26 will be used to ensure that the results obtained will be more valid. According to Schumacker & Lomax discovered the following checklists in developing SEM research, which will be detailed below:

a. Theoretical Model & Data Preparation

Structural Equation Modeling (SEM) study starts by outlining the rationale and purpose of the study, followed by establishing a strong theoretical basis. In this study, SEM is utilized to examine the relationships between the key variables. This study involves four latent variables, which are unobservable and require measurable indicators for analysis. The exogenous latent variable is Merdeka Curriculum, with four indicators: Flexible Learning (X_1), Essential Materials (X_2), Diverse Teaching (X_3), and Technology Integration (X_4). Teacher Creativity is the first endogenous latent variable, measured by Fluency (Y_1), Flexibility (Y_2), and Elaboration (Y_3). The second endogenous latent variable is Professional Development, with indicators including Teacher

Professional Education (PPG) (Y_4), Training (Y_5), and *Guru Penggerak* (Y_6). Lastly, Teacher Burnout is measured by Workloads (Y_7), Administrative Tasks (Y_8), and Student Behavior (Y_9). The rationale for the study is to understand how these factors interrelate in the educational context.

b. Model Specification

Model specification involves determining every relationship and parameter in the model that is of interest to the researcher.[120] The goal is to create a theoretical model that accurately represents the sample data. If the model is incorrectly specified, it can lead to incorrect estimates of the parameters, which is known as specification error.

After the theoretical framework model is constructed, it is then transformed into a path diagram to depict the causal relationships between exogenous variables and endogenous variables before finally being illustrated as a full model of the structural model.

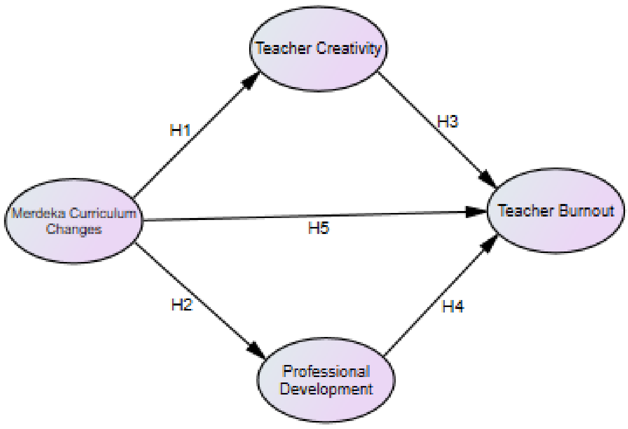

Figure 3. 4. Path Diagram for the Model Specification

According to Ghozali[121], there are two essential steps in structural modeling: constructing the structural model by linking latent constructs, both endogenous and exogenous, and specifying the model by linking endogenous or exogenous constructs with indicator or manifest variables. The full model of structural model for this study can be seen as follows:

[120] Randall E. Schumacker and Richard G. Lomax, *A Beginner's Guide to Structural Equation Modeling: Third Edition* (Routledge, 2012), 214.
[121] Ghozali, *Model Persamaan Struktural, Konsep Dan Aplikasi Dengan Program AMOS 24, 30*.

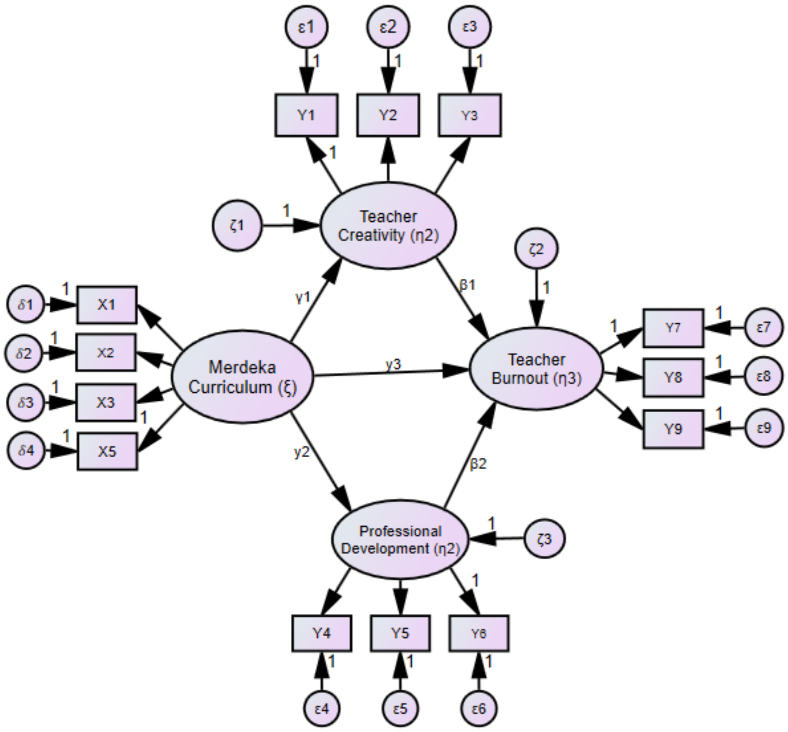

Figure 3. 5. Path diagram of the Structural Model Analysis

Explanation of the Structural Model Analysis Diagram:

a) In this structural model analysis above, there is one exogenous latent variable, symbolized by ξ "ksi", which represents Merdeka Curriculum.

b) This exogenous variable is measured by manifest variables, symbolized by X, in this diagram there are four exogenous manifest variables.

c) The error value associated with the latent variable X is symbolized by δ "delta".

d) The endogenous variables in this structural model are symbolized by η "eta", where η1 = teacher creativity, η2 = professional development, and η3 = teacher burnout.

e) These endogenous variables are measured by manifest variables, symbolized by Y.

f) The error value associated with the latent variable Y is symbolized by ε "epsilon".

g) All endogenous latent variables are assigned a *structural error* or *error term*, symbolized by ζ "zeta".

h) The regression relationship between exogenous and endogenous constructs is symbolized by γ "gamma" and is represented with a single arrow →.

i) The regression relationship between one endogenous construct and another endogenous construct is symbolized by β "beta".

j) The measurement model for X:

   - Exogenous Latent Variable $\xi 1$
     $X1 = \lambda 1.1\ \xi 1 + \delta 1$
     $X2 = \lambda 2.1\ \xi 1 + \delta 2$
     $X3 = \lambda 3.1\ \xi 1 + \delta 3$
     $X4 = \lambda 4.1\ \xi 1 + \delta 4$

k) The measurement model for Y:

   - Endogenous Latent Variable $\eta 1$
     $Y1 = \lambda 1.1\ \eta 1 + \varepsilon 1$
     $Y2 = \lambda 2.1\ \eta 1 + \varepsilon 2$
     $Y3 = \lambda 3.1\ \eta 1 + \varepsilon 3$
   - Endogenous Latent Variable $\eta 2$
     $Y4 = \lambda 5.1\ \eta 2 + \varepsilon 4$
     $Y5 = \lambda 4.1\ \eta 2 + \varepsilon 5$
     $Y6 = \lambda 6.1\ \eta 2 + \varepsilon 6$
   - Endogenous Latent Variable $\eta 3$
     $Y7 = \lambda 7.1\ \eta 1 + \varepsilon 7$
     $Y8 = \lambda 8.1\ \eta 1 + \varepsilon 8$
     $Y9 = \lambda 9.1\ \eta 1 + \varepsilon 9$

c. Model Identification

Model identification in Structural Equation Modeling (SEM) is the process of checking whether the model can be properly estimated.[122] It involves counting the number of parameters to be estimated and comparing them with the amount of information available from the data, such as variances and covariances. Based on this comparison, a model can be classified into three types: *underidentified* (not enough information to estimate the model), *just-identified* (the number of parameters equals the available information), and *overidentified* (more information than parameters,

[122] Schumacker and Lomax, *A Beginner's Guide to Structural Equation Modeling*, 2012.

which allows both estimation and model testing). Ideally, a model should be overidentified so that it can be properly tested and evaluated.

d. Model Estimation

In model estimation, selecting the appropriate technique is crucial for estimating the parameters in both the measurement and structural models. These estimates represent the population parameters based on the sample data. For example, if the data meets specific criteria such as having no missing data, outliers, and following a normal distribution, *maximum likelihood estimation* (ML) may be suitable. ML estimates, along with standard errors and chi-square tests, are appropriate for interval-scaled variables that are normally distributed.

According to Bollen[123], MLE has several key features and works best with large samples because its properties become more accurate as the sample size increases. In the estimation stage of the SEM model, the output provides detailed information that helps researchers understand the relationships between variables. The key results include regression weights, which show the strength and direction of the relationship between variables; direct effects, which indicate the immediate influence one variable has on another; and indirect effects, which reflect the influence of one variable on another through one or more mediating variables. Together, these results help explain how the variables in the model are connected and whether the hypotheses proposed in the study are supported by the data.

e. Model Testing

In this step, the model's adequacy is evaluated through an examination of various Goodness-of-Fit criteria. Prerequisite testing is used to determine if the structural model meets the assumptions required by SEM, using the AMOS 26 application to establish the model's adequacy based on specific goodness-of-fit criteria. According to Ghozali[124], there are three types of goodness-of-fit measures: absolute fit measures, incremental fit measures and parsimonious fit

[123] Bollen, *Structural Equations with Latent Variables*.
[124] Ghozali, *Model Persamaan Struktural, Konsep Dan Aplikasi Dengan Program AMOS 24, 63-66*.

measures. This study will use the first type, which is absolute fit measures, because they can assess the overall fit of both the structural and measurement models together. Some tests of model adequacy using absolute fit measures are as follows:

1. *Likelihood Chi-square Test:*

   The Likelihood chi-square test is used to develop and test if the model being examined fits the estimated model. A high Chi-square value relative to the degree of freedom indicates that the observed covariance or correlation matrix differs significantly from the predicted matrix, resulting in a probability level smaller than the significance level. Conversely, a low Chi-square value will result in a probability level greater than the significance level, indicating that the observed and predicted covariance or correlation matrices do not differ significantly. Empirical data is considered a good fit to the theory or model, or highly significant, if the Chi-square probability value is greater than the significance level (α) and the Chi-square value is less than or equal to the critical value ($\chi^2\alpha;df$). In other words, a small Chi-square value and a high p-value ($p > 0.05$) suggest that the model fits the data well.

   However, it is important to note that the Chi-square statistic is highly sensitive to sample size[125]. In large samples, even minor differences between the model and the data can result in a statistically significant Chi-square value ($p < 0.05$), making it appear that the model does not fit well, despite the differences being trivial. Therefore, researchers are advised to interpret the Chi-square value alongside other model fit indices such as CFI, TLI, RMSEA, and SRMR for a more comprehensive evaluation.

2. *CMIN/DF Test:*

   The CMIN/DF test is the chi-square value divided by the degree of freedom. Some studies recommend using this ratio as a measure of fit.

[125] Kline, *Principles and Practice of Structural Equation Modeling*. 269

According to Wheaton et al., a ratio value of 5 or less is considered a reasonable fit.[126]

3. *Root Mean Square Error of Approximation (RMSEA):*

This test is a measure that attempts to improve the tendency of the chi-square statistic to reject the model with a large sample size. If RMSEA $\leq$ 0.05, it indicates a Close Fit. An RMSEA between 0.05 and 0.08 suggests a Margin Fit, and an RMSEA $\geq$ 0.09 indicates a Poor Fit.

4. *Goodness of Fit Index (GFI):*

The Goodness of Fit Index (GFI), developed by Jöreskog and Sörbom, is a non-statistical measure used to evaluate how well a model fits the empirical data.[127] If the GFI is $\geq$ 0.90, it indicates a Good Fit. A GFI between 0.80 and 0.90 suggests a Margin Fit, while a GFI $<$ 0.80 indicates a Poor Fit.

5. *Tucker Lewis Index (TLI):*

The Tucker Lewis Index, also known as the non-normed fit Index (NNFI), combines parsimony into the comparison index between the proposed model and the null model. TLI is an incremental fit index that compares the tested model with the baseline. A TLI $\geq$ 0.90 is considered a Good Fit, $0.80 \leq$ TLI $<$ 0.90 indicates a Margin Fit, and a TLI $<$ 0.80 means the model is Not Fit.

6. *Normed Fit Index (NFI):*

The Normed Fit Index is a measure of comparison between the proposed model and the null model. If the NFI is $\geq$ 0.90, it indicates a Good Fit. An NFI between 0.80 and 0.90 suggests a Margin Fit, while an NFI $<$ 0.80 indicates a Not Fit model.

7. *Comparative Fit Index (CFI):*

The Comparative Fit Index is an incremental fit index. This index is relatively insensitive to sample size and is less influenced by model complexity. If the CFI is $\geq$ 0.90, it indicates a Good Fit. A CFI between

[126] Ghozali, 64.
[127] Ghozali, 65.

0.80 and 0.90 suggests a Margin Fit, while a CFI < 0.80 indicates a Not Fit model.

f. Model Modification

In SEM-based research, if the model's estimation results show a poor Goodness of Fit (GoF), re-specification is required. Modifying the model will affect the initial estimation results. The goal of this process is to improve the overall model fit, such as reducing the chi-square value. Model modification involves changing the initial estimated model to improve fit, particularly the GoF index. However, this should be done carefully and aligned with the theory. For example, covariances may be added between the error terms of related latent variables. Any changes should be supported by a strong theoretical foundation. Without clear theoretical justification, modifications might improve statistical fit but can weaken the model's theoretical meaning and validity.