CHAPTER III

RESEARCH METHOD

3.1 Population and Sample

Population is not only the amount of subject or object being studied, but also includes the characteristics of subject or object. It is a generalization term that contains specific qualities of subject or object determined by the researcher to learn and then draw a conclusion (Supardi, 1993). In this research, the population used are person taxpayers and body taxpayers listed in *KPP Purworejo*. The number of taxpayers registered in KPP Purworejo as of December 31, 2018 are as follows:

Body Taxpayer	2,555
Person Taxpayer (Employee)	42,012
Person Taxpayer (Non-employee)	8,831
Total	53,398

Source: Division of data and information processing, 2019

Sample is part of the population that is the subject of study as "representative" of the population members (Supardi, 1993). Thus, the researcher takes not all qualified taxpayers but only a portion of them to be used as respondent. Hair, Black, Babin, & Anderson (2010) suggested for regression analysis, it is recommended that 15 to 20 observations for each independent variables. Therefore, the sample used in this study is about 60 taxpayers consisting of person taxpayer and body taxpayer. The researcher will take 40 respondents as person taxpayer and 20 respondents as body taxpayer.

The sampling method in this study is using convenience sampling method. Convenience sampling is the most commonly method used in quantitative studies in which the researcher is free in determining subjects as the representative of the population. Generally, the researcher tends to select subjects that are easily accessible or readily available. In other words, all qualified subjects does not have the same opportunity to participate in the study (Etikan, Musa, & Alkassim, 2016).

3.2 Sources and Data Collection Method

The data were obtained by distributing questionnaires to qualified respondents. Questionnaires are written questions that will be filled out by respondents in the form of writing, checking, or circling responses (Harmon, Morgan, & Harmon, 2001). This research was using questionnaires that is measured by an interval scale with four possible answers, which are Strongly Disagree (SD), Disagree (D), Agree (A), and Strongly Agree (SA).

The questions were derived from previous studies that have been conducted by other researchers and are relevant to this study. Thus, the questions might be considered as valid since it has been already tested and used effectively in prior researches.

3.3 Research Variables

The variables in this research are dependent and independent variables. The dependent variable is taxpayer satisfaction with e-filing system while the independent variables are information quality, system quality, and service quality.

3.3.1 Taxpayer Satisfaction

Taxpayer satisfaction is a response which describes the conformity of users' expectation to the result of the overall system applied. It is important

to measure the level of taxpayer satisfaction as an indicator that system being implemented by the government is successful (Ningrum & Andi, 2016).

The questionnaire for this variable consists of five questions, of which one question is adopted from Zaidi, Henderson, & Gupta (2017), two questions from Teo et al. (2009), and two questions from Veeramootoo, Nunkoo, & Dwivedi (2018).

3.3.2 Information Quality

Information quality according to Chang et al. (2005) cited in Munusamy, Run, Chelliah, & Annamalah (2012) is defined as the level at which the information provided can meet the user needs. It represents the users' perception of the output quality in the form of information generated by the system used which covers informativeness, relevancy, timeliness, and accuracy.

The questionnaire for this variable consists of eight questions, of which seven questions are adopted from Chen (2010) and one question from Saha, Nath, & Salehi-Sangari (2012).

3.3.3 System Quality

System quality is the level at which e-filing website is easy to use and can provide quick access to users. In other words, the system enables to complete immediately the orders given by the users and does not crash. The quality of system is considered as a success factor that affect user satisfaction with the e-filing system (Widyadinata & Toly, 2014).

The questionnaire for this variable consists of eleven questions, of which four questions are adopted from Chen (2010), two questions from Saha et al. (2012), two question from Zaidi et al. (2017), two questions from Chen et al. (2015), and one question from Veeramootoo et al. (2018).

3.3.4 Service Quality

Service quality is the measurement of how well the service provided in meeting the user needs and is able to fulfil users' expectation. It can be assessed from employees' attitude, condition of the office, and facilities being offered. A good service quality will satisfy users, thus they tend to give favorable feedbacks (Permatasari, Susilo, & Topowijono, 2015).

The questionnaire for this variable consists of seven questions, of which six questions are adopted from Chen (2010) and one question from Zaidi et al. (2017).

3.4 Analytical Techniques

3.4.1 Validity Test and Reliability Test

Validity test is used to measure the accuracy of the questionnaire. A questionnaire is considered valid if the questions in the questionnaire reflect the overall content being tested. Thus, the researcher must analyze whether the questions represent the entire content or not in order to ensure that the questionnaire is valid. The validity of the questionnaire can be measured statistically by using a software (Matondang, 2009).

Validity test is done by comparing the coefficient of r count and r table. If r count $\geq$ r table and have positive value, it can be concluded that the questionnaire is valid. On the contrary, if r count < r table, it means that the

questionnaire is invalid. This test measures the validity of questionnaire as a whole, not the validity of each question contained in it (Matondang, 2009).

Reliability test is used to measure the extent to which the results of a test can be trusted. The results can be trusted if in several times of the measurement on the same subject, the results are relatively constant, with the condition that the aspects measured on the subject have not changed (Matondang, 2009).

In the reliability test, the number of Cronbach Alpha is used to estimate whether the response of respondents is reliable or not. The response is considered reliable if the scores obtained are consistent and the number of Cronbach Alpha > 0.70. Otherwise, if the number of Cronbach Alpha ≤ 0.70, it means that the response of respondents is not reliable. In addition, a negative number might be resulted in the test, which means that something is wrong with the data and the researcher needs to do a review (Ghozali, 2013).

3.4.2 Descriptive Statistical Analysis

Descriptive analysis is an analysis conducted to assess the characteristics of the data derived from the sample. The purpose of descriptive statistics is to provide an overview of the data seen from the value of average, standard deviation, variance, maximum, and minimum. By conducting this analysis, the readers will find it easier to understand the data in more detail.

3.4.3 Classical Assumption

Classical assumption is a test that must be fulfilled in a linear regression model, thus the model can be categorized valid as an estimator. A

good regression model meets the classical assumptions of normality, multicollinearity, and heteroscedasticity.

1. Normality Test

   Normality test ensures that the data are normally and independently distributed. With normal distribution of the data, it implies that the data can be considered to represent the population. Normality test of the data can be statistically done by comparing the value of Kolmogorov-Smirnov Z and Z table. The value of Kolmogorov-Smirnov Z can be calculated by performing nonparametric test (Ghasemi & Zahediasl, 2012). If the value of Kolmogorov-Smirnov Z < Z table, then the distribution is classified normal. On the contrary, if the Kolmogorov-Smirnov Z > Z table, the distribution is classified not normal.

2. Multicollinearity Test

   Multicollinearity test is intended to find out whether independent variables in the regression model are correlated or not. A good regression model should not have correlation between independent variables, which means there is no multicollinearity. Correlated independent variables are termed as not orthogonal because the value of correlation not equal to zero (Ghozali, 2013). Detection of the multicollinearity presence or absence in the regression model can be seen from the value of tolerance and variance inflation factor (VIF) with equation as follow:

   $$\text{VIF} = \frac{1}{Tolerance}$$

The equation above describe that the low value of tolerance leads to the high value of VIF. Tolerance measures the variability of selected independent variables that are not explained by other independent variables. Thus, the value of tolerance > 0.10 or equal to VIF < 10 indicates the absence of multicollinearity. Meanwhile, the value of tolerance < 0.10 or equal to VIF > 10 indicates the presence of multicollinearity (Ghozali, 2013).

3. Heteroscedasticity Test

Heteroscedasticity test aims to test whether there is inequality of variance from residual in one observation to another observation. Heteroscedasticity occurs when the variance of residual in one observation is different from another observation, while homoscedasticity occurs when the variance of residual in one observation is the similar with another observation. However, a good regression model should be free from heteroscedasticity (Ghozali, 2013). There are several ways to detect the presence or absence of heteroscedasticity, one of them is using Glejser test. Glejser proposes to regress an absolute value of residual to the independent variable. According to Ghozali (2013), with a significant level of 5%, heteroscedasticity can be detected as follow:

a. If P-value $\leq 5\%$, it means that there is indication of heteroscedasticity.

b. If P-value $> 5\%$, it means that there is no indication of heteroscedasticity.

3.4.4 Multiple Regression

Multiple regression analysis is used to know the influence of independent variables on dependent variable expressed in the form of mathematical equations. In this study, the equation of multiple regression can be formulated as follow:

$$TP = \alpha + \beta IQ + \beta SYSQ + \beta SERVQ + \varepsilon$$

Where:

TP = Taxpayer Satisfaction

α = Constanta

β = Coefficient Regression

IQ = Information Quality

SYSQ = System Quality

SERVQ = Service Quality

ε = Residual Error

3.4.5 Hypothesis Testing

Hypothesis testing is performed to prove the influence of independent variables toward dependent variable. In this study, the researcher will test the coefficient of determination (R^2), simultaneous regression test (F Test), and partial regression test (T Test).

1. Coefficient of Determination (R^2)

Coefficient of determination (R^2) is interpreted as a contribution of influence given by independent variables on the dependent variable. The value of determination coefficient is used to know the extent to which the

independent variables can influence the dependent variable with condition that the result of f test in the regression analysis is significant. Otherwise, if the result of f test in the regression analysis is insignificant, then the value of determination coefficient cannot predict the influence of independent variables to the dependent variable. The R^2 value exists between 0 and 1 ($0 < R^2 < 1$). However, the use of R^2 is biased against the number of independent variables which were entered into the model. Therefore, it is recommended to use adjusted R^2 in determining the best regression model (Ghozali, 2013). The value of adjusted R^2 might be negative or positive. The greater R^2 (closes to 1), the better result for the regression model. Meanwhile, the closer value to 0, it means the independent variables as a whole cannot explain its influence toward the dependent variable.

2. Simultaneous Regression Test (F Test)

Simultaneous regression test aims to determine whether there is simultaneous influence among the independent variables on the dependent variable. In this study, the researcher wants to test whether the independent variables of information quality, system quality, and service quality simultaneously influence the dependent variable of taxpayer satisfaction. According to Ghozali (2013), the use of 5% significance value will obtain the following results:

a. If P-value $\leq$ 5%, then H_a is accepted. It means the independent variables simultaneously influence the dependent variable.

b. If P-value $> 5\%$, then H_a is rejected. It means the independent

variables do not simultaneously influence the dependent variable.

3. Partial Regression Test (T Test)

The researcher uses partial regression test (t test) to partially test the influence of independent variables toward dependent variable. The purpose of this test is to find out whether each of independent variables is significantly influence dependent variable. In this study, the researcher wants to test whether the independent variables of information quality, system quality, and service quality partially significantly influence the dependent variable of taxpayer satisfaction. By using a significance level of 5%, the result of t test is concluded based on the following criteria:

a. If P-value $\leq 5\%$, then H_a is accepted. It means the independent variable

partially significantly influence the dependent variable.

b. If P-value $> 5\%$, then H_a is rejected. It means the independent variable

does not partially significantly influence the dependent variable.

4. Discussion

The researcher will compare the findings with the theoretical review, previous researches, and hypotheses about the influences of information quality, system quality, and service quality on taxpayer satisfaction with e-filing system. The result will determine whether there is consistency with theories and empirical findings in the previous research. If there is any difference, the researcher will find out the reason behind it.