Application of The Rasch Model in The Development of Dimension of The Measurement of Tax Fairness Perception

Meita Larasati, Cotoro Mukri
Universitas Muhammadiyah Prof. Dr. HAMKA
email: meita.larasati@uhamka.ac.id

Keywords: fairness, exchange fairness, horizontal fairness, vertical fairness, retributive fairness, policy fairness, administrative fairness

ABSTRACT
This study aims to investigate the measurement instrument of the perception dimension of tax fairness using the Rasch model method. The dimensions of the perception of tax fairness are divided into seven dimensions, namely general fairness, exchange fairness, horizontal fairness, vertical fairness, retributive fairness, policy fairness, and administrative fairness. The sampling technique employed purposive sampling by classifying several criteria for respondents. The distribution of data used Google docs to as many as 122 individuals. The results of this study indicate that the Cronbach alpha value (KR-20) is 0.71. The question item with code PF2 which is included in the personal fairness dimension with a logit value of +0.91 is the most difficult item for respondents to answer. There are three question items that are considered misfit because the logit value of the item is greater than the sum of the mean and standard deviation (1.27). The three items are VF 1 (1.72), RF1 (1.48), and EF2 (1.29). The contributions of this study are: (1) to enrich the literature on tax fairness perceptions by using the Rasch model as a data analysis tool, and (2) empirically, to validate the measurement instrument in the perception of tax fairness dimensions which will be useful for further research.
INTRODUCTION

Most countries in the world consider taxes as the main source of state income. The government considers that taxes are an important component of state revenue as a means of government financing (Tabandeh, Jusoh, Nor Ghani, & Zaidi, 2013). Government systems in almost all over the world have always maximized their tax collection systems to increase state revenues in order to finance budgets effectively and efficiently (Maseko, 2014). Therefore, tax collection policies must consider the behavior and perceptions of taxpayers regarding the fairness of the tax collection system. Taxes are defined as payments made by taxpayers to support state income. The tax system must be fair to all citizens. When the tax system is deemed unfair, taxpayers will think that the system is bad and inappropriate.

Most of the previous research related to tax compliance has focused on economic factors, such as profitability, assets, R&D costs, operating costs, leverage, and financial reports (Frank, Lynch, & Rego, 2009). Recently, the approach of non-economic factors related to tax compliance has been highlighted by various groups. This problem began to be seen from various perspectives on the behavior of taxpayers, including the perception of tax fairness.

Perceptions of tax fairness arise from the events or thoughts of a person comparing himself to others (van den Bos, Peters, Bobocel, & Ybema, 2006). The perception of fairness is considered very important because it affects a person’s willingness to see a rule as something that can be trusted and can encourage cooperative behavior (Lind, 2001). Several previous studies related to the perception of tax fairness have always been based on theories, models, and actions experienced by taxpayers and organizations in accordance with the tax fairness framework (Doherty & Wolak, 2012; Konow, 2001). Theoretical understanding of the perception of tax fairness is formed from contextual factors in the formulation of a framework for the perception of justice and a more specific understanding of the framework of tax fairness.

Public economic theory states that the tax system can be evaluated fairly through vertical and horizontal fairness. Vertical fairness means that taxpayers with different incomes must be subject to tax burdens at different levels, namely taxpayers with higher income will get higher taxation and will automatically pay higher taxes than individuals with lower income. Horizontal fairness is defined as equal treatment of individuals who have the same conditions. Taxpayers with equal economic conditions will get an equal tax burden (Jurney, Rupert, & Wartick, 2017). However, the two dimensions of fairness are still controversial in relation to progressive and proportional taxes. Therefore, the tax fairness dimension begins to develop and becomes an interesting issue to continue to be researched.

Several other dimensions related to the perception of tax fairness have been investigated by previous research, namely (Bobek, Hageman, & Kelliher, 2013) about procedural fairness and policy fairness. Procedural Fairness deals with the results of tax collections by the government which are distributed fairly for the welfare of the community. Policy fairness contains the justice of laws or policies that govern the taxation system in a country. Another dimension of the perception of tax fairness is exchange fairness which is a reflection of the exchange of tax contributions and the benefits obtained from the government for citizens (Gilligan & Richardson, 2005). This dimension of fairness can present the perception of fairness from taxpayers about the tax system if the benefits obtained from the government are proportional to the tax contributions they sacrifice. In addition, several other dimensions of tax fairness that have been investigated by previous research include a preference for progressive or proportional taxation, namely personal fairness, tax rate fairness, and general fairness (Gilligan & Richardson, 2005).

The development of the perceptual dimension of tax fairness makes these measurement instruments worthy of validation with appropriate analytical tools to test non-parametric social research. As far as the researchers’ best knowledge, there has been no research validating the measurement instrument of the perception of tax fairness using the Rasch model. Researchers use the Rasch model because the test tool is very suitable for investigating social research because it will produce information that meets the definition of measurement (Bond, T. & Fox, C., 2015). The purpose of this study was to validate measurement instruments on seven dimensions of tax fairness perceptions, namely general fairness, exchange fairness, horizontal fairness, vertical fairness, retributive fairness, policy fairness, and administrative fairness.

General fairness measures an individual’s assessment of the fairness of the income tax system.
Exchange fairness relates to reciprocity between taxpayers and the government. Horizontal fairness treats tax imposition equally among taxpayers in the same financial situation. Vertical fairness is assessed based on an individual's ability to pay taxes and a preference for a tax rate structure, either a fixed or progressive rate. Retributive fairness relates to penalties imposed on taxpayers for their negligence. Personal fairness relates to individual interests, while administrative fairness relates to tax laws or policies and procedures used by tax authorities.

This study makes several contributions, first to enrich the literature on tax fairness perceptions by using the Rasch model as a data analysis tool. Second, empirically this study can validate the measurement instrument in the perception of tax fairness dimensions which will be useful for further research.

Rasch Model in the Development of Measurement Tools

Classical theory (Classical Test Theory / CTT) is the most commonly used approach to research in the social sciences and psychology (Wibisono, 1992). In social science research, the difficulty in measuring is to determine quantitative weighting of latent quantitative phenomena (Cavanagh & Waugh, 2011). So far, research using the CTT approach believes that the pure score (T) and error (E) are formulations to produce visible scores (X). Error becomes the basis of situational conditions out of control (Sumintono & Widhiarso, 2013).

Nowadays, social and psychological measurement tools are developing rapidly with reference to CTT, but several criticisms and resistance have begun to develop regarding the CTT approach, one of which is (Alagumalai, Curtis, & Hungi, 2005) states that the sample, visible scores, and pure scores affect the item difficulty level, test difficulty level, and error assumption for all respondents. Affirmed by Michell (2002) that the type of data obtained from questions on opinions and attitudes is nominal or ordinal so that an appropriate analytical tool is needed to carry out measurements. This criticism presents the Item Response Theory (IRT) on improvements from CTT. The Rasch model is a form of application of the Item Response Theory.

The Rasch model has the ability to predict hilling data (missing data) based on individual response patterns, therefore the statistical results of the Rasch model are considered more accurate (Sumintono & Widhiarso, 2013). Compared to other methods, the Rasch model is considered to be superior in social and psychological research. In addition, the Rasch model can produce a standard error measurement score on the instrument used so as to increase the accuracy of the calculation. The social and psychological research approach, especially in non-parametric data, the Rasch model is able to adjust the data to its natural conditions, which is a cotinum for the characteristics of quantitative data, whereas CTT is considered incapable (Sumintono & Widhiarso, 2013). Ordinal data can be transformed into ratios through the Rasch model which refers to the principle of probability so that the level of data accuracy will be better. Rasch model in analyzing data will adjust to the model whereas in CTT the model is formed based on the available data. Therefore, the Rasch model will validate the instrument to produce more holistic information and meet measurement definitions (Bond, T. & Fox, C., 2015). The Rasch model emphasizes five important parts in the analysis using the Rasch model, namely the calibration and estimation ability of items, item characteristic curves in parameter models, item and instrument information functions, interaction maps between items and respondents, and items and respondents which is fit / misfit.

RESEARCH METHOD

This research employed a quantitative study with data analysis using the Rasch model assisted by Winstep software. Rasch model was used in this study because it is in accordance with the research objectives, namely to validate the taxpayer's justice instrument. The Rasch model is considered capable of seeing the interaction between respondents and items simultaneously. The assessment of respondent data was seen from the logit value which could reflect the probability of selecting an item in a group of respondents.

The method used a survey by distributing questionnaires to respondents in accordance with the objectives of this study. The questionnaire in this study was distributed using Google docs. The respondents collected in this study were 122 respondents. The sample collection technique in this study used purposive sampling by determining the sample based on certain criteria in accordance with the research objectives. The samples chosen...
in this study were Indonesian citizens who were taxpayers, taxpayers who regularly reported and paid income tax every period and were at least 20 years old. The sample criteria were selected based on several considerations. First, an individual who was 20 years old was considered of sufficient age to be taxpayers and able to generate income. Second, according to the researcher's consideration, taxpayers who routinely reported and paid taxes have a direct perception of tax fairness. The variable in this study was the perception of tax fairness which included seven dimensions. These dimensions were general fairness, exchange fairness, horizontal fairness, vertical fairness, retributive fairness, policy fairness, and administrative fairness. There were 20 question items given to respondents to measure perceptions of tax fairness. This study adapted the tax fairness instrument from previous research (Smulders, 2013) which refers to (Gilligan & Richardson, 2005). The research instrument distributed to respondents has been adjusted to the existing conditions in Indonesia.

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Statement</th>
<th>Item Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>General Fairness</td>
<td>I believe everyone pays the appropriate income tax rate under the current income tax system.</td>
<td>GF1</td>
</tr>
<tr>
<td></td>
<td>I believe that the government uses the right amount of tax revenue to achieve social goals.</td>
<td>GF2</td>
</tr>
<tr>
<td></td>
<td>I think the government is using too much of its tax revenues for unnecessary welfare assistance.</td>
<td>GF3</td>
</tr>
<tr>
<td>Exchange Fairness</td>
<td>I receive a suitable reciprocal from the government in exchange for my income tax payments.</td>
<td>EF1</td>
</tr>
<tr>
<td></td>
<td>I think it is fair if those with low incomes receive more benefits from the government than those with high incomes.</td>
<td>EF2</td>
</tr>
<tr>
<td></td>
<td>The high amount of income tax that I have to pay is in accordance with the benefits I receive from the government.</td>
<td>EF3</td>
</tr>
<tr>
<td>Horizontal Fairness</td>
<td>I think it's fair that several people who have the same amount of income will pay the same amount of income tax.</td>
<td>HF1</td>
</tr>
<tr>
<td></td>
<td>I think it is fair to have to pay the same amount of income tax as other taxpayers who have the same income as me.</td>
<td>HF2</td>
</tr>
<tr>
<td></td>
<td>In my opinion, it is fair if every taxpayer who has the same income has the same income tax rate.</td>
<td>HF3</td>
</tr>
<tr>
<td>Vertical Fairness</td>
<td>I think it's fair that those with high incomes will be subject to higher tax rates than those on lower incomes.</td>
<td>VF1</td>
</tr>
<tr>
<td></td>
<td>I think it's fair that those with middle income are taxed less than those on high incomes.</td>
<td>VF2</td>
</tr>
<tr>
<td></td>
<td>The income tax rate paid by high-income people is exorbitant.</td>
<td>VF3</td>
</tr>
<tr>
<td>Retributive Fairness</td>
<td>I think it is fair if taxpayers who deliberately do not pay taxes are punished with the same penalty burden, regardless of the amount of tax not paid.</td>
<td>RF1</td>
</tr>
<tr>
<td></td>
<td>I think to be fair, the penalty rate for tax evasion must depend on the level of tax non-compliance.</td>
<td>RF2</td>
</tr>
<tr>
<td>Policy Fairness</td>
<td>I believe that I pay a fair tax rate under the current income tax system.</td>
<td>PF1</td>
</tr>
<tr>
<td></td>
<td>Compared to other taxpayers, I pay more income tax rates.</td>
<td>PF2</td>
</tr>
<tr>
<td></td>
<td>I think those with middle income pay a fair income tax rate according to the current income tax system.</td>
<td>PF3</td>
</tr>
<tr>
<td>Administrative Fairness</td>
<td>There are several ways available to correct a mistake in calculating my tax liability rate, at no additional cost.</td>
<td>AF1</td>
</tr>
<tr>
<td></td>
<td>The system administration applies consistent income tax throughout the year for taxpayers.</td>
<td>AF2</td>
</tr>
</tbody>
</table>

Participants were asked to evaluate statements on five Likert scales (1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree) by filling in the fields provided. The scoring results in this study are politomic.

**RESEARCH RESULT**

Data sourced from research respondents tabulated in Ms. Excel software and then converted and analyzed the data using Winstep Rasch. The
consideration of using the software is adjusted to the research objectives.

Instrument Reliability Test

The reliability test of the instrument with the Rasch model in this study is shown in Table 2. In the table, it is informed that the amount of data from 122 respondents and 20 question items related to the perception of tax fairness dimensions is 2440 and produces a Chi Square value of 6194.3612 and has a value of degrees of freedom (d.f) amounting to 6175 (p=0.000 and p<0.01).

The reliability analysis of this instrument produces two types of output, first to translate the results of the analysis from the respondent (person) and second to explain the items. The table of respondents explains whether or not the respondents analyzed in this study. Meanwhile, the item table describes whether or not an item of measurement is fit. Table 2 stating the person measure value of 0.59 which shows that the respondent has a relatively high perceived tax fairness score. This means that most of the respondents answered “agree” to the instrument item measuring the perception of tax fairness. Separation value is 1.39. The value of separation shows the quality of the instrument with respondents and items because it can identify groups of respondents and groups of items (Sumintono & Widhiarso, 2013). The strata separation uses the following equation:

\[ H = \left[ \frac{4 \times 1.39}{3} \right] + 1 = 2.2 \]

Based on this equation, the value of \( H = 2.2 \) is obtained which is rounded to 2. This means that the respondents are divided into two large groups, namely groups with high and low perceived fairness values. A similar equation was applied to tabulate the item strata, so that the value \( H = 6.32 \) was obtained and rounded into 6 groups. In accordance with this equation, the items are divided into 6 groups based on the level of difficulty to be approved by the respondent.

Table 2. Summary Statistic
Interactions between respondents and items were assessed with Cronbach's alpha (KR-20). Table 2 shows that the Cronbach alpha value (KR-20) is 0.71. This value is included in the good category (α > 0.70) and means that there is an agreement between the respondent and the instrument used. While the reliability value for the item is 0.95 which shows a very good value (α > 0.94) (Sumintono & Widhiarso, 2013). Based on the results obtained, the data is stated in accordance with the criteria and requirements of the Rasch model so that the analysis steps can be continued.

The item group is divided into eight strata by dividing the logit value of the item into 6 equal groups. The logit value of the item can reflect the respondent's assessment of the item to be more objective because the raw data that is ordinal has been converted into ratio data so that it meets the integer criteria (Sumintono & Widhiarso, 2013).

**Item Value Test**

Table 3 in this research describes the level of difficulty of the items answered by the research respondents. Measure (logit item value) sorted from the most difficult item to approve (highest value) to the easiest item to approve (lowest value) (Sumintono & Widhiarso, 2013). In addition, the table provides information about the logits for each item.

**Fit Order Item Value Test**

The fit order item value test is carried out to provide information on whether an item is fit or not (Sumintono & Widhiarso, 2013). The items are sorted from least suitable (misfit) to most suitable (fit). The fit and misfit criteria are obtained by comparing the INFIT MNSQ value of each item in the table with the sum of the mean and standard deviation values. A larger logit value indicates the item is in a misfit.

<table>
<thead>
<tr>
<th>Item Measure</th>
<th>Statistics: Measure Order</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Table 3. Item Measure</strong></td>
<td></td>
</tr>
</tbody>
</table>
Based on the table above, the sum of the average value and standard deviation (1.02 + 0.25) is 1.27. Referring to this value, it can be seen that there are three question items that are considered misfit because the logit value of the item is greater than the sum of the mean and standard deviation (1.27). The three items are VF 1 (1.72), RF1 (1.48), and EF2 (1.29).

**Person / Respondent Value Test**

Table 5 provides information about the logit for each respondent in this study. This table indicates that respondents have the highest perception of tax fairness compared to other respondents (Sumintono & Widhiarso, 2013). Respondents indicated that they tend to answer more strongly agree and agree to the tax fairness perception questionnaire.

From the table above, it can be seen that the respondent with serial number 049 and female has the highest logit value than the others, which is 1.86. This indicates that respondent number 049 has the highest fairness perception of the taxation system in Indonesia compared to other respondents. Meanwhile, the male respondent number 091 had the lowest logit value than the others, which was -0.47. Respondents -0.47 have the lowest perception of tax fairness compared to other respondents.

Total Count with a value of 20 for all respondents shows that each respondent in this study answered all questions on the questionnaire given by the researcher. The total of all question items is 20 items. Therefore, no data is lost.
### Tabel 5. Person Measure

<table>
<thead>
<tr>
<th>Person</th>
<th>REAL SEP.</th>
<th>REL.</th>
<th>Item REAL SEP.</th>
<th>REL.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.39</td>
<td>0.66</td>
<td>4.49</td>
<td>0.95</td>
</tr>
</tbody>
</table>
Unidimensionality Test

The instrument’s unidimensionality test is conducted to determine the ability of the instrument to measure what should be measured (Sumintono & Widhiarso, 2013), in this research is the perception of tax fairness. The unidimensionality test has the same function as the instrument validity test. The minimum limit percentage of the unidimensionality value is 20%, this means that the instrument is fulfilled. A raw variance value of more than 40% means better and 60% means special. Another thing to note is that the unexplained variance should ideally not exceed 15%.

Table 6. Unidimensionality

| Raw Variance Explained by Measures | 7.8286 28.1% | 28.5% |
| Raw Variance Explained by Persons  | 1.7370 6.2%  | 6.3%  |
| Raw Variance Explained by Items   | 6.0916 21.9% | 22.2% |
| Raw Unexplained Variance (Total)  | 0.0000 71.9% | 71.5% |

Table 6 shows the raw variance value of 28.5%, which means that it meets the unidimensionality requirements. Besides that, the value of the unexpected variance does not reach 15%, this means that all the unidimensionality requirements have been fulfilled properly.

Rating Scale Test

This test is conducted to verify whether the rating used in this study can be understood well by the respondent or not. The Rasch model is able to identify respondents’ assumptions about the rating of the questions presented in the instrument (Sumintono & Widhiarso, 2013). This study uses five scale choices in the form of a Likert rating for each question item. The range of choices is strongly disagree, disagree, neutral, agree, and strongly agree.

Table 7 shows that the mean value of observations (observation average) starts at logit 0.01 for the choice of score 1 (strongly disagree), logit 0.02 (disagree), logit 0.29 (neutral), logit 0.73 (agree), and logit 1.07 (very agree). The increase in logit scores between scores 1 and 2 was very small, only 0.01. This indicates that the respondent is not very responsive to the difference between scale 1 (strongly disagree) and scale 2 (disagree). Meanwhile, between scales 3, 4 and 5 have a logit value that moves quite significantly. This indicates that respondents are quite responsive to differences in the ranking scale.

In addition, the Andrich Threshold value must be considered to test the appropriateness of the politomical values used in this study. This value moves from NONE to negative and leads to positive. This value shows that the options used are valid for the respondent. The non-consecutive Andrich Threshold values mean that the choice of options is better to simplify.
CONCLUSION

The purpose of this study was to validate measurement instruments on seven dimensions of tax fairness perceptions, namely general fairness, exchange fairness, horizontal fairness, vertical fairness, retributive fairness, policy fairness, and administrative fairness. The contribution of this study, first, to enrich the literature on tax fairness perceptions by using the Rasch model as a data analysis tool; second, empirically this study can validate the measurement instrument in the perception of tax fairness dimensions which will be useful for further research.

The results of this study indicate that the Cronbach alpha value (KR-20) is 0.71. This value is included in the good category (α > 0.70) and means that there is an agreement between the respondent and the instrument used. While the reliability value for the item is 0.95 which shows a very good value (α > 0.94). The most difficult question item for respondents to agree is on the item code PF2 which is included in the personal fairness dimension with a logit item value of +0.91. There are three question items that are considered misfit because the logit value of the item is greater than the sum of the mean and standard deviation (1.27). The three items are VF 1 (1.72), RF1 (1.48), and EF2 (1.29).

Respondent 049 who is female has the highest logit value, which is 1.86. This shows that these respondents have a high perception of tax fairness in Indonesia. Meanwhile, the lowest logit value was owned by the male respondent 091, namely -0.47. Respondents -0.47 have the lowest perception of tax fairness compared to other respondents. The rating scale of the question instruments in this study has insignificant difference values on a scale of 1 and 2 with a logit value of 0.01 and 0.02. Meanwhile, between scales 3, 4 and 5 have a logit value that moves quite significantly. This indicates that respondents are quite responsive to differences in the ranking scale.

The limitation of this research lies in the number of samples that are still insufficient to represent the perceptions of taxpayers in Indonesia. Future studies are expected to enrich the study sample and examine the differences between the tax perception dimensions more clearly.


