Journal of Indonesian Economy and Business
Volume 32, Number 3, 2017, 234 – 246

THE IMPACT OF A DISAGGREGATION STRATEGY WITH SUPPLEMENTARY DISPLAYS TOWARDS BALANCED SCORECARD PERFORMANCE EVALUATION

Frisky Jeremy Kasingku
Faculty of Economics, Universitas Klabat, Indonesia
(frisky.kasingku@gmail.com)

Goedono[1]
Faculty of Economics and Business, Universitas Gadjah Mada, Indonesia
(gudono@feb.ugm.ac.id)

ABSTRACT

Balanced Scorecard has become a popular management tool around the world. Despite its benefits, the balanced scorecard creates a bias called the common- measures bias. The bias associated with a balanced scorecard should be eliminated so that the optimal benefits of the balanced scorecard can be obtained. To eliminate the bias, a disaggregation strategy is suggested. This disaggregation strategy is found to be a mitigation strategy to solve the common-measures bias. However, there is a small amount of empirical evidence about the quality of the decisions produced by using a disaggregation strategy. Furthermore, to increase the decisions' quality, an information display was found to be helpful. Therefore, this study aimed to investigate the effect of a disaggregation strategy, in the context of a balanced scorecard, toward the decisions' quality regarding the balanced scorecard's performance evaluation with a different information display. This study used an experimental method with the design of 3x2x2 between the subject's factorial designs. The results indicated that decision makers with supplementary tabular and graphic displays would exhibit a greater judgment consensus than decision makers who were given traditional separate displays. Moreover, those who received the supplementary table displays exhibit a greater judgment consensus and consistency than those who were given a graphical display. In conclusion, a disaggregation strategy with table and graphic displays could improve a judgment's consensus, a traditional display can improve a judgment's consistency, and a table display could exhibit greater judgmental consensus and consistency than a graphic display. This study contributed theoretically and practically.

Keywords: *balanced scorecard, decision quality, disaggregation strategy, information display*

INTRODUCTION

The balanced scorecard has become one of the most popular management tool around the world (Rigby & Bilodeau, 2015). The balanced scorecard was originally developed as a performance measurement tool, but the function of this management tool has developed. The balanced scorecard has transformed into a strategic management tool that connects the vision, mission, goals, and strategy of a company; it communicates the strategy to the lowest level of the organization (Kaplan & Norton, 1992; Kaplan & Norton, 1993; Kaplan & Norton 1996).

The use of the balanced scorecard as a performance measurement tool brings several problems (Lipe & Salterio, 2000; Itner, Larcker, & Meyer, 2003; Banker, Chang, and Pizzini, 2004; Ding & Beaulieu, 2011). The previous research discusses the complexity of using a balanced scorecard. It generates a lot of information that exceeds a human's capacity to process information. Consequently, the excess of information leads to a phenomenon called information overload (Ding & Beaulieu, 2011). If the company cannot cope or find ways to

[1] The late Goedono

mitigate this information overload, the investment in a balanced scorecard could be futile since its users could not obtain the optimal benefits of using it (Neumann, Roberts, & Cauvin, 2010).

Lipe and Salterio (2000) uncover the impact of information overload on a balanced scorecard's performance evaluation. They conducted an experiment to examine the judgments made by the participants about common and unique measures. They argue that managers should give equal weight to both measures, since both measures describe the strategy employed by each business unit. Furthermore, there are psychological theories that are used to support their arguments. As a result, the participants tend to pay more attention to the common measures than the unique measures. They concluded that the information overload creates a common-measures bias, where managers tend to focus on the common measures rather than the unique measures. Furthermore, a lot of researchers were struggling to find a way to mitigate the common-measures bias such as Libby, Salterio, and Webb (2004), Dilla and Steinbart (2005a), and Roberts, Albright, and Hibbets (2004). In addition, Utami (2001) explains that the common-measures bias has occurred in performance evaluations in Indonesia. She experimented with 116 students as the participants for her study; as a result, she concluded that cognitive biases such as the common-measures bias do not only occur in America, but also in Indonesia.

Roberts, Albright, and Hibbets (2004) found mitigation for the common-measures bias. They examined the disaggregation strategy in evaluating both the common and unique measures. They argue that with assistance, the quality of the decisions can be improved. The assistance given in their study was a disaggregation strategy. Furthermore, through the experiments, the participants were given assistance by weighting each common and unique measure. In addition, the participants gave their judgment through a scoring mechanism for each measure, based on the given weight. After that, they make a decision based on the overall score. As a result, when the disaggregation strategy was implemented, the participants focused on both the common and unique measures, therefore, the common-measures bias had been successfully mitigated.

Research conducted by Roberts et al. (2004) has found mitigation for the common-measures bias, but there are some gaps associated with this research. Their study did not examine the quality of the decisions resulting from the disaggregation strategy. The quality of the decisions can be measured by the level of consensus, consistency, and easiness in evaluating the performance of the balanced scorecard. (Dilla & Steinbart, 2005b). The consensus is the degree of agreement between each decision maker (Ashton, 1985), while consistency is a provision in the act. In the context of measuring performance, consistency is seen as a provision in the performance evaluation and the giving of bonuses (Dilla & Steinbart, 2005b).

Referring to the research conducted by Dilla and Steinbart (2005b), a supplementary display of tables and graphs can increase the level of consensus, consistency, and easiness in evaluating performance. Tabular displays facilitate users to directly compare the performance of each unit (Fennema & Kleinmuntz, 1995; Kleintmuntz & Schkade, 1993). Organizing information in a tabular form induces managers to use a higher proportion of their attributes in the decision-making process than when the information is presented in a separate display (Dilla & Steinbart, 2005b). Therefore, by providing additional information in tabular form, this will increase the likelihood of an evaluator finding an optimal decision maker.

Furthermore, Martinson, Davison, and Tse (1999) found that graphical displays also facilitate decision-making in the complexity of the balanced scorecard. Managers' judgments can also be enhanced through providing information in a graphical display (Dilla & Steinbart, 2005b). In addition, tables and graphs can be useful as additional tools to help managers so that they can produce a quality

assessment (Dilla & Steinbart, 2005b). The use of information displays can reduce the complexity of allocating a bonus. On the other hand, previous studies on the comparison of information displays (graphs vs. tables) have not found consistent results.

Based on the research gaps that have been disclosed previously, the research problem is formulated as follows: "Does a disaggregation strategy, along with a supplementary information display, affect the balanced scorecard's performance evaluation?" This study examines the effect of a disaggregation strategy with a supplementary display towards the level of consensus, consistency, and easiness in evaluating the performance of the balanced scorecard. Furthermore, this study also examines the comparative use of tables and graphics in evaluating performance.

This research benefits researchers and practitioners. For researchers, this study extends the balanced scorecard's literature by adding and confirming the research by Roberts et al. (2004). Furthermore, this research benefits researchers in the accounting information field. This study adds information on display literature by examining the levels of consensus, consistency, and easiness in evaluating balanced scorecards' performances. Furthermore, this research is useful to practitioners when evaluating performance. The use of the disaggregation strategy with an information display is expected to reduce the common-measures bias and improve the consensus, consistency, and easiness in evaluating performance.

LITTERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

1. Balanced Scorecard

Kaplan and Norton (1992) introduced a performance measurement tool known as the balanced scorecard. At that time, the balanced scorecard functioned as a tool for development. Furthermore, there are some aspects that were introduced in their first publication. First, the balanced scorecard is a powerful tool for measuring performance, using financial and non-financial measures. Moreover, there are four perspectives to the balanced scorecard when it was first introduced (financial, customer, internal business, innovation and learning). Lastly, the balanced scorecard is a tool that predicts long-term performance; therefore, it has a futuristic orientation. Furthermore, in a subsequent publication, Kaplan and Norton (1993) continued their explanations regarding the performance measurement tool. The balanced scorecard does not only function as a measurement tool, but also as a motivating tool for management systems so that satisfactory progress can be achieved.

2. Bounded Rationality and Information Overload

Rationality is the process of decision making, based on logical thinking. It considers the risk preferences and value of the judgment that leads to optimal decision making. Therefore, it can be concluded that an individual assessment will be limited to each individual's rationality. Furthermore, there are several factors that affect the rationality of individuals, such as a lack of information, time constraints and costs, the lack of individual memory capacity, and the individual's intelligence. Thus, these factors become a barrier for individuals to make optimal decisions (Bazerman & Moore, 2009).

Bazerman and Moore (2009) explain the concept of bounded rationality. They indicate that an individual's assessment tends to deviate from their rationality. Tversky and Kahneman (1974) study in some depth about this bounded rationality. Consequently, heuristics provides an efficient way to simplify complex decisions. In fact, it is undeniable that the individual is faced with a lot of information when decision-making, that's why people often avoid making decisions. Individuals who have to make decisions based on the information overload situation tend to filter out information. Unfortunately, important information could be filtered through that kind of decision making (Bazerman & Moore, 2009).

Eppler and Mengis (2004) define information overload as the reception of too much information. The amount of information received by an individual is positively related to

their decision-making performance and the quality of the decision. However, the positive relationship only lasts until a certain point. When the information reaches that certain point, the decision-making performance, and the quality of the decision, will decrease. In addition, the information provided beyond that point will not be considered in the decision-making process, therefore, the phenomenon called information overload emerges. The phenomenon of information overload creates a burden for the individual, it affects the ability of decision makers to set priorities and makes it more difficult for an individual to process information (O'Reilly, 1980; Eppler & Mengis, 2004).

There are various causes of information overload. Those causes are changes in organizational design, the characteristics of the information, the capacity of individuals to manage information, the tasks and processes used in local decision-making, and information technology (Eppler & Mengis, 2004). Baldacchino, Armistead and Parker (2002) suggested several solutions to prevent information overload. First, they suggest learn more, remember more, use analytical tools for decision making, as well as to relax when making decisions. Furthermore, the use of software helps to filter information, consequently, it can help individuals in making decisions when faced with a lot of information. The last strategy suggested is the goal-oriented approach. This approach aims to focus the company's energy into areas that will bring it success. Several studies have been conducted to test the effects of information overload, that research was conducted by Swain and Haka (2000) and Hwang and Lin (1999).

3. Balanced Scorecard and Information Overload

Ding and Beaulieu (2011) argue that the information generated from the use of the balanced scorecard is classified as an information overload. Referring to the description and Mengis and Eppler (2004), we could conclude that the information generated by the balanced scorecard will positively affect the performance of decision-making, as well as the quality of the decision, until at a certain point, the information provided will degrade the accuracy of the decision-making.

Lipe and Salterio (2000) explain that the information overload occurs in the balanced scorecard's performance evaluation. The information overload is generated by a lot of measures in a balanced scorecard. These measures make managers simplify their decision-making processes. They found that managers tend to pay more attention to the common measures rather than both the common and unique measures. The phenomenon became known as the common-measures bias. Thus, it can be concluded that with the use of the balanced scorecard, a lot of information is generated and it creates an information overload (Lipe & Salterio, 2000).

4. Disaggregation Strategy

Having discovered the common-measures bias in the assessment of a balanced scorecard's performance, some researchers have attempted to examine the variables that might mitigate the common-measures bias. One of them is Roberts et al. (2004). They identify a mitigating factor for the bias. Roberts et al. (2004) examined the role of a disaggregation strategy in mitigating the bias in the common-measures bias. Lipe and Salterio (2000) provide that the bias could be caused by the lack of an individual's cognitive effort incurred when making decisions. Furthermore, Roberts et. al (2004) argued that by doing a disaggregation, the information can be broken down into various dimensions so that individuals could give more cognitive efforts.

Roberts et al. (2004) continues that these dimensions will be assessed separately by the individual. When individuals assess one dimension rather than assessing multiple dimensions at the same time, their attention will be focused on one dimension so that short-term memory people will be relieved of the burden of thinking about the other dimensions. Processing this information will make people give an adequate cognitive effort, so that people will

focus on all the dimensions. As a result, the common-measures bias could be mitigated through that strategy. Consequently, the strategy will lower and raise the burden of the task at the same time. For example, the use of these strategies will reduce the amount of information that must be processed by the individual, but the amount of effort and time required to process the information will increase as the number of evaluations and calculations increases (Roberts et al., 2004).

5. Human Information Processing

One topic of research in the field of accounting is the study of the management of information by humans. Researchers examined the behavior that occurs in the management process. Libby and Lewis (1982) provide a model of the information management process. They explain that one of the variables contained in the input is the method of presentation. The method is divided into three categories: Time format (numeric, graphic, verbal), sequence, and the aggregation and disaggregation of data. Furthermore, the method of presentation affects the assessment process. It produces an output that varies according to the given input.

6. Human Information Processing and Balanced Scorecard

Some studies examine the effects of information management in the context of the balanced scorecard (Lipe & Salterio, 2000; Itner, Larcker, & Meyer; Banker, Chang, and Pizzini, 2004; Ding & Beaulieu, 2011). These studies tend to view the information's management processes such as subjectivity, the mood difference, as well as an individual's cognitive effort. Moreover, Dilla and Steinbart (2005b) studied the effect of inputs given to the individual's judgment in evaluating performance.

Dilla and Steinbart (2005b) examine the effect of additional comparative views towards the consensus, consistency, and the perceived easiness of evaluation. The input given in this study is information displays. They divide the information displays into two categories (tables and graphs). As a result, the consensus and consistency of the evaluators are higher when they are given extra help, compared to those who did not get the information displays.

7. Hypothesis Development

7.1. Consensus

Lipe and Salterio (2000) responded that the common-measures bias can occur when an individual gives less cognitive effort. To facilitate the individual in providing a cognitive effort, Roberts et al. (2004) offer a disaggregation strategy, so that, with the disaggregation strategy, people will make an adequate effort, so that people will pay attention to the common and unique measures. On the other hand, Roberts et al. (2004) acknowledge that the disaggregation strategy can increase the time it takes to process information and can improve the individual's cognitive effort. The amount of time required for decision-making will affect the performance and quality of the decisions (Eppler & Menger, 2004). Consequently, Kleinmuntz and Schkade (1993) and Ashton (1985) have argued that it needs help to make more accurate decisions. Furthermore, Ashton (1985) proposed that accuracy is strongly associated with consensus; therefore, consensus can be used as a substitute for accuracy, even though there are some limitations to the changeover. Moreover, he argued that the provision of information, and consensus in decision making, can be improved.

Libby and Lewis (1982) and Schaubroeck and Muralidhar (1991) explain that presentation methods can affect the accuracy of decisions. Furthermore, Dilla and Steinbart (2005b) have argued that the graphic and tabular formats can help managers to increase their consensus decisions. Schaubroeck and Muralidhar (1991) explain that the graphs can help managers to improve the speed of their decision making, while tables can help managers to make decisions more comprehensive.

Therefore, it can be summarized that because there are limitations in the use of the disaggregation strategy, the assistance of graphs and tables will improve the decision-making's performance in terms of consensus (Roberts et

al., 2004; Kleinmuntz & Schkade, 1993; Schaubroeck & Muralidhar, 1991; Ashton 1985; Dilla & Steinbart, 2005b). Through the description above, it can be hypothesized that:

H1a: Decision makers who received the supplementary tabular display will exhibit a greater judgment consensus than decision makers who are given a traditional separate display.

H1b: Decision makers who received the supplementary graphic display will exhibit a greater judgment consensus than decision makers who are given a traditional separate display.

7.2. Consistency

The balanced scorecard was originally used as a performance measurement tool (Kaplan & Norton, 1992, 1993). This performance measurement tool subsequently evolved into a strategic management tool for strategy implementation and evaluation (Kaplan & Norton, 1996). The balanced scorecard with the function of a strategic management tool would be useful if it is connected to the compensation (Kaplan & Norton, 2001; Niven, 2002). However, an inconsistent relationship occurs between performance evaluation and compensation given (Ittner, Larcker, & Meyer, 2003). Therefore, the consistency between these relationships is an important point, in that consistency is an important indicator in assessing the quality of decisions (Dilla & Steinbart, 2005b).

A disaggregation strategy can mitigate the common-measures bias, but because there are some obstacles, such as the amount of time required, the consistency of the decision is questioned (Roberts et al, 2004). Furthermore, Dilla and Steinbart (2005b) have argued that by giving additional supplementary displays (tables and graphs), the quality of the decision, such as its consistency, can be improved. Therefore, it can be hypothesized that:

H2a: Decision makers who received the supplementary tabular display will exhibit a greater judgment consistency than decision makers who are given a traditional separate display.

H2b: Decision makers who received the supplementary graphic display will exhibit a greater judgment consistency than decision makers who are given a traditional separate display.

7.3. Task Easiness

There are some limitations to the use of a disaggregation strategy (Roberts et al., 2004). The use of a disaggregation strategy does not necessarily improve the quality of the decisions, because individuals must spend more time and effort making them. Consequently, Dilla and Steinbart (2005b) proposed aids to improve the quality of the decisions. They argue that because a display is designed to help facilitate comparisons between units, decision makers can do their job more easily. Therefore, it can be hypothesized that:

H3a: Decision makers who received the supplementary tabular display will rate the performance evaluation and bonus allocation decisions as being easier than decision makers who are given a traditional separate display.

H3b: Decision makers who received the supplementary graphic display will rate the performance evaluation and bonus allocation decision as being easier than decision makers who are given a traditional separate display.

7.4. Cognitive Fit Theory

The use of the information display helps decision-makers to improve the quality of the decisions they make (Libby and Lewis, 1982). Furthermore, there are research studies that test and compare the information displays (Vessey, 1991; Kleinmuntz & Schkade, 1993; Jarvenpaa, 1990; Umanath & Vessey, 1994; Schaubroeck & Muralidhar, 1991; Dilla & Steinbart, 2005). These studies compared the use of graphs and tables. As a result, the cognitive theory explains that both displays will fit into the needed tasks. The graph is more suitable for tasks that require

the identification and understanding of the relationship as well as for comparisons.

On the other hand, the table is useful for tasks that require individuals to extract values and then combine all of these values into an overall assessment (Dilla & Steinbart, 2005b). Therefore, because of the strategy, a disaggregation strategy requires those assessments for the overall assessment, then a table is more suitable for this strategy (Dilla & Steinbart, 2005; Roberts et al., 2005b). Through the previous description, it can be hypothesized that:

H4a: Decision makers who received the supplementary tabular display will exhibit a greater judgment consensus than decision makers who are given a graphical display.

H4b: Decision makers who received the supplementary tabular display will exhibit a greater judgment consistency than decision makers who are given a graphical display.

H4c: Decision makers who received the supplementary tabular display will rate the performance evaluation and bonus allocation decisions as easier to make than decision makers who are given a graphical display.

RESEARCH METHOD

1. Research Design

This research used experimental methods with 3x2x2 between the subject's full-factorial designs. The first variable used is a presentation format. The format of the presentation is divided into three levels, namely a tabular display, a graphic display, and one without the help of a display. Furthermore, the performance of both divisions in the common measures is divided into two levels. One division will have a better performance in the common measures and vice versa. The next factor is the performance between the subjects in both divisions in the unique measures, which is divided into two levels. One division will have a better performance in the unique measures and vice versa.

Each participant will randomly get one of the 12 treatments. The participants will get one view of the information (tables, graphs, or traditional) with one common measure of the performance patterns and one unique measure of the performance patterns. This study employs the experimental instruments used by Lipe and Salterio (2000), which will be modified according to the experimental context. The modifications will be made referring to the instrument developed by Roberts et al. (2004) and Dilla and Steinbart (2005).

2. Participants

This study uses students as the study's participants, with the assumption that they have sufficient knowledge about the balanced scorecard's measures. The use of students is considered suitable because the researcher only wanted to examine the cognitive processes of the individual when making a general decision (Nahartyo & Utami, 2016). The cognitive processes of each individual are generally the same, so the use of students as the subjects of this study is considered as adequate to describe the cognitive processes that are similar to the cognitive processes experienced by individuals who evaluate performance. The use of students as research subjects is in line with previous research conducted by Lipe and Salterio (2000), Dilla and Steinbart (2005a; 2005b), and Roberts et al (2004).

This study uses student who have or are taking courses in Management Accounting. Furthermore, there are 12 cells in the experiment that will be carried out, for that, assuming a minimum of 10 participants for each cell, the required number of participants in this study is at least 120 participants. The use of a minimum of 10 participants is in accordance with the recommendation of Nahartyo and Utami (2016).

One hundred and ninety two students participated in the experiment. A total of 40 participants failed to qualify from the manipulation test, so that data from 152 participants are eligible for this study. Of the participants used, 44.08% (67 participants) are male, 55.92% (85 participants) are women.

Furthermore, 40.79% (62 participants) are MBA students, while 59.21% (90 participants) are accounting students.

3. Dependent Variable

3.1. Consensus

The consensus variable is measured by using a pairwise consensus (Dilla & Steinbart, 2005b). A pairwise consensus shows the consistency of any decision taken by each participant in a single cell. There are two steps in determining the pairwise consensus. First, the pairwise consensus for individuals is determined. Participants are given a value corresponding to the decision taken. For example, if a participant decides to give a higher value for the teen fashion division, and the other participants gave higher ratings for the division with work clothes, then the given value is -1. Furthermore, if two participants gave similar assessments in one division, then the value assigned is 1. Moreover, if one of the participants assesses one division higher than the other division, while the other participants assess both divisions the same, then the value given is 0. The consensus for individuals can range between -1 and 1.

Secondly, the average value of the individual consensus is calculated to assess the overall consensus in one cell. For example, if there are four participants in one cell where three participants assess the teen fashion divisions higher than the division with work clothes then the calculation is as follows. To calculate the value of the participants who rate the teen fashion division higher: ((3*1) + (3*-1)) / 4 = 0.5. Furthermore, in order to calculate the value of the participants who rate the work clothes division higher: ((3*-1) + (3*1)) / 4 = -0.5. Lastly, the overall value of consensus in one cell is calculated as: ((3*0.5) + (1*-0.5)) / 4 = 0:25. Thus, the value of consensus for a single cell is 0.25. The average value of the consensus within a cell can range between 0 and 1. The average value of consensus is the value of the individual consensus on a single cell.

3.2. Consistency

Consistency is an appropriate assessment of the performance and bonuses. To assess the consistency, a code is given based on the results of the assessment and the awarding of bonuses to each individual. There are two categories for this dependent variable, which are consistent and inconsistent. In addition, researchers will give the code 0 for inconsistent results while the code 1 is for consistent results. After giving the code to the participants who are consistent and inconsistent, the proportion of consistency is calculated using data from the participants who are consistent. The proportion of consistency is used to measure the consistency for each participant.

3.3. Task Easiness

Task easiness is measured using a questionnaire that has been used by Dilla and Steinbart (2005b). The questionnaire was administered using a Likert-type scale ranging from -5 (strongly disagree) to +5 (strongly agree). Participants who agree with the statements of the questionnaire indicate a low level of difficulty.

3.4. Task and Procedures

The participants were given the role of a senior executive who has a duty to assist the company's financial director to evaluate the performance of the two divisions of the Modern Women's Clothing company. The company specializes in women's clothing. They were then given a case that describes the mission of the company. The case would then provide a balanced scorecard for each division. Each Balanced scorecard has 16 measures, 4 measures for each perspective. Eight measures illustrate the common measures and the remaining illustrates the unique measures.

The participants are given instruments randomly. This randomization of the instruments is done through several stages. First, the researchers print all the instruments that are used, in accordance with the targeted sample. Second, all of the instruments that have been printed are scrambled by the researchers. Third, the researcher is aided by the research assistants to distribute the instruments that have been randomized in advance, so that the instruments

become more randomized, so that the goal of randomization could be achieved.

Information displays are given differently in each case. Furthermore, the participants were asked to write a composite score for each division. After that, the participants were asked to nominate who is more suitable for promotion, and then they were asked to allocate a bonus of 50 million rupiah for the two division managers.

After answering the three questions, the participants were asked to answer a few questions about the ease of the performance evaluation. They were asked to give feedback on a scale of -5 (strongly disagree) to +5 (strongly agree). The same scale is used to answer the question of manipulation. At the final stage of the experiment, the participants were asked to provide their demographic data.

3.5. Pilot Test

A pilot test was performed, in order to test whether the case that will be presented can be understood by the subjects. Furthermore, it was performed to understand any shortcomings and errors in the research instrument. Therefore, the researcher conducted a pilot test on 32 accounting sciences postgraduate students from Gadjah Mada University. There were 29 participants who passed the manipulation. Improvement of the instruments was carried out based on the suggestions obtained from the pilot test.

3.6. Statistical Tools

ANOVA is used as the statistical tool to analyze the data in this study. ANOVA is used when researchers test the asymmetrical relationship between independent variables that are measured with a non-metric scale, and the dependent variable which is measured by a ratio or interval scale (Gudono, 2014). A chi-square test (McClave, Benson& Sincich, 2000; Bolboaca, Jantschi, Sestras, Sestras, & Pamfil, 2011) and an ANOVA were conducted to see the effect given by a number of factors that could affect the study's results.

RESULTS AND DISCUSSION

1. Manipulation Test

A manipulation test was conducted to determine the participants' understanding of the material in the case. Furthermore, there are four questions for the participants to measure their understanding of the case materials. The first question is "Division 1 and 2 have the same target market". The second question is "Some measures in the teenage clothing division and the work clothing division are different". The third question is "It is appropriate if the measures in Divisions 1 and 2 are different". The fourth question is "Financial measures are as important as the other perspective measures". These questions also measure the understanding of the participants toward a balanced scorecard. They are asked to give their answers based on the scale of -5 (strongly disagree) to +5 (strongly agree). Understanding of these experiments can be seen if a participant successfully answers at least three out of four manipulation question.

2. Hypothesis Test Result

The first hypothesis test result is presented in Table 1. These results indicate that the significant value of information to the consensus is 0.000, it is smaller than the value α = 0:05 so that a post-hoc test was undertaken. The post-hoc test results showed that the average value of the tables (0.2026) is higher than the average value of the traditional display, while the average value of the graphs (0.1180) is higher than the average value of the traditional display. Therefore, Hypotheses 1a and 1b are supported.

Table 1. Hypotheses 1 - 3 Test Result

Variable	df	Mean Square	F	Sig
Information Display H1	2	0.621	5.878E+34	0.000
Information Display H2	2	0.271	3.533	0.033
Information Display H3	2	2.117	0.432	0.650

Source: Primary Data (2016)

The second hypothesis test result is presented in Table 1. These results indicate that the significant value of information to the consistency is 0.033, it is smaller than the value

α = 0:05 so that a post-hoc test was done. This post-hoc test results showed that the average value of the tables (-0.84907) is lower than the average value of the traditional display, while the average value of the graphs (-0.165003) is lower than the average value of the traditional display. Thus, Hypotheses 2a and 2b are not supported.

The third hypothesis test result is presented in Table 1. These results indicate that the significant value of information to the task easiness is 0.0650, it is higher than the value of α = 0:05 so that the post-hoc analysis are not tested. Therefore, there is no significant difference in using the information display with the easiness of the task, so that Hypotheses 3a and 3b are not supported.

The fourth hypothesis is analyzed using Hypotheses 1-3's post hoc results. If the average value of the tables is higher than the average value of the graphs, or the traditional graph, then Hypotheses 4a, 4b and 4c are supported. The results show that the average value of the tables (0.0847) is higher than the average value of the graphs for the consensus variables, thus Hypothesis 4a is supported. Furthermore, the average value of the tables (0.80096) is higher than the average value of the graphs for the consistency variable, so that this supports Hypothesis 4b. The average value of the tables (-0.0890) is lower than the average value of the graphs for the task easiness task, so Hypothesis 4c is not supported.

3. Discussion

After performing the statistical tests, here are some of the findings from this study. First, the provision of information displays affects the level of consensus among the participants. The various information display aids produce different levels of consensus. The results of this study show that the presentation of information in the form of tables resulted in the highest degree of consensus, followed by graphical displays and traditional displays respectively. The results of this study did not show it fully supported the previous research conducted by Dila and Steinbart (2005b). They mentioned that their research did not find clarity. Therefore, the results of this study confirm that the provision of information in the form of tables and graphs increases the consensus, compared to the traditional type of display.

Second, the provision of information displays affects the level of consistency of the performance evaluations and bonuses. The provision of information in a traditional display generates the highest level of consistency, followed by tables and then displays using graphs, respectively. Furthermore, this finding confirms the results of the research conducted by Dila and Steinbart (2005b). The provision of information, either in the form of traditional displays, graphs, or tables, improves consistency, but the provision of information in the form of traditional displays provides the highest consistency among all the information displays.

Furthermore, the third result of this study shows that providing information does not affect the degree of easiness in evaluating performance. In fact, there is no significant difference shown by the provision of information displays. The latest results of this study indicate that the provision of information in the form of tables generates a higher level of consensus and consistency than the provision of information in the form of graphs. On the other hand, a traditional display makes it easier to evaluate performance, compared to the provision of information in tables and graphs.

The third and fourth research results are in line with the research conducted by Dila and Steinbart (2005b). First, there is no significant difference between the provision of information and the ease in evaluating performance. Results are also affected by the age factor. Secondly, according to the theory of cognitive fit (Vessey, 1991; Umanath & Vessey, 1994), tables and graphs have their respective functions for different tasks. Tables will be more beneficial if used for symbolic tasks and graphs for spatial tasks. According to Dila and Steinbart (2005b), the tasks of evaluating performance and the bonus payments are symbolic and spatial, but are dominated by the symbolic task. Therefore, in

accordance with the theory of cognitive fit, the results of this study can explain in detail why tables show levels of consensus and higher consistency compared to displays using graphs.

CONCLUSION, LIMITATIONS, AND RECCOMENDATIONS

1. Conclusion

A balanced scorecard can bring many benefits to its users when the bias that can reduce these benefits is eliminated. A disaggregation strategy successfully mitigates the common measures bias. Although it can mitigate the common measures bias, the quality of the resulting decisions through the disaggregation strategy has not been studied. Based on the analysis of the research, there are a number of conclusions for this study, namely: Provision of information in the form of tables, graphs, and using traditional methods will increase the consensus. The provision of information in the form of tables produces a higher consensus compared to information provided by graphs and traditional displays. The provision of information in the form of tables, graphs, and traditional displays is able to increase the consistency. A traditional display is the display that improves the consistency the most, followed by displays featuring tables and graphs. Information displayed in the form of tables, graphs, or in a traditional manner do not affect the task's easiness in evaluating performance. Information displays in the form of tables generate a higher degree of consensus and consistency than displays using graphs do. A disaggregation strategy used with information displays affects the quality of the resulting decisions (consensus, consistency, and task easiness), depending on which information display is used.

2. Implications

The results of this study have several implications, both theoretically and practically. First, the results showed that the disaggregation strategy, when used with information displays, can improve the consensus. The results of this study add references to the consensus and consistency literature relating to evaluating performance on the basis of a balanced scorecard. Consistency is one of the important factors in the evaluation of performance. Giving bonuses to managers should be proportional to the value of a given performance evaluation. Therefore, the use of traditional information displays can provide the highest proportion of consistency. Next, the results of this study add to the literature on tables vs graphs; although this has been studied for some considerable time, researchers have not yet reached a level of consensus in determining the most appropriate type of information display. Therefore, this study adds to the literature asserting that tables are more suitable for symbolic tasks and graphs for spatial tasks.

Furthermore, on the practical side, this study contributes to improving the quality of decisions produced when evaluating a balanced scorecard's performance. Managers can use an information display, in the form of tables, to improve the consensus, and traditional displays to improve the consistency.

3. Limitations

This study has several limitations. First, the implementation of the experiment was paper based. Along with the times, some software has been developed to support the ease of Balanced scorecard-based performance evaluations. One of them is the QPR scorecard. The software allows the user to design a Balanced scorecard in accordance with the will of its users. Therefore, the implementation of a computerized experiment will be more effective. Although it can be more effective, the use of a paper based experiment was able to generalize in the same way a computerized experiment can be generalized (Dila & Steinbart, 2005b). Second, the use of students as the research's participants. Although previous research mentions that the use of participants who have knowledge of the Balanced scorecard is enough, the use of participants who have experience in the field of Balanced scorecard will produce more tangible results. Hence because of the difficulty in finding participants who have experience of Balanced scorecard, the use of participants who

have knowledge of it is sufficient for this study. Third, the age factor affecting the task easiness variable is not explored in this study.

4. Recommendations

Suggestions for further research based on several of the limitations of this study are: Further research may retest the quality of the decisions resulting from the disaggregation strategy for the benefit of external validity. Further research can use a computerized system to conduct the experiments. Subsequent research may consider the use of participants who have experience in evaluating the performance-based Balanced scorecard. Further research can better explore the extraneous variables that can affect the results of this study, such as age, Grade Point Average (GPA), background, and more.

REFERENCES

Ashton, A., 1985. "Does Consensus Imply Accuracy in Accounting Studies of Decision Making?" *The Accounting Review*, 173-185.

Baldacchino, K., C. Armistead, and D. Parker, 2002. "Information Overload: It's Time to Face the Problem." *Management Services,* 18-19.

Banker, R., G. Potter, and D. Srinivasan, 2000. "An Empirical Investigation of an Incentive Plan That Includes Nonfinancial Performance Measures." *The Accounting Review*, 65-92.

Bazerman, M., and D. Moore, 2009. *Judgment in Managerial Decision Making*. New Caledonia: John Wiley & Sons, Inc.

Bolboaca, S., L. Jantschi, A. Setras, R. Setras, and D. Pamfil, 2011. Pearson-Fisher Chi-Square Statistic Revisited. *Information,* 528-545.

Dilla, W., and P. Steinbart, 2005a. "Relative Weighting of Common and Unique Balanced Scorecard Measures by Knowledgeable Decision Makers." *Behavioral Research in Accounting*, 43-55.

Dilla, W., and P. Steinbart, 2005b. "The Effects of Alternative Supplementary Display Formats on Balanced Scorecard Judgments." *International Journal of Accounting Information System*, 159-176.

Ding, S., and P. Beaulieu, 2011. "The Role of Financial Incentives in Balanced Scorecard-Based Performance Evaluations: Correcting Mood Congruency Bias." *Journal of Accounting Research, 1223-1247*.

Eppler, M., and J. Mengis, 2004. "The Concept of Information Overload: A review of Literature from Organization Science, Accounting, Marketing, MIS, and Related Disciplines." *The Information Society*, 325-344.

Fennema, M., 1993. Anticipations of Effort and Accuracy in Multiattribute Choice. Disertasi. Universitas Illinois.

Gudono, 2014. *Analisis Data Multivariat*. Yogyakarta: BPFE.

Hoque, Z., 2014. "20 Years of Studies on the Balanced Scorecard: Trends, Accomplishments, Gaps and Opportunities for Future Research." *The British Accounting Review*, 33-59.

Hwang, M., and J. Lin, 1998. "Information Dimension, Information Overload and Decision Quality." *Journal of Information Science*, 213-218.

Itner, C., D. Larcker, and M. Meyer, 2003. "Subjectivity and the Weighting of Performance Measures: Evidence from a Balanced Scorecard." *The Accounting Review*, 725-758.

Jarvenpa, S., 1990. "Graphic Displays in Decision Making – the Visual Salience Effect." *Journal of Behavioral Decision Making*, 247-262.

Kaplan, R., and D. Norton, 1992. "The Balanced Scorecard-Measures that Drives Performance." *Harvard Business Review*, 107-121.

Kaplan, R., and D. Norton, 1993. "Putting the Balanced Scorecard to Work." *Harvard Business Review,* 134-147.

Kaplan, R., and D. Norton, 1996. *The Balanced Scorecard: Translating Strategy into Action*. Boston: Harvard Business School Press.

Kaplan, R., and D. Norton, 2001. *The Strategy-Focused Organization: How Balanced Scorecard Companies Thrive in the New Business Environment*. Boston: Harvard Business Press.

Klentmuntz, D., and D. Schkade, 1993. "Information Display and Decision Processes." *Psychological Science*, 221-227.

Libby, R., and B. Lewis, 1982. "Human Information Processing Research in Accounting: The State of the Art in 1982." *Accounting, Organizations and Society*, 231-285.

Libby, T., S. Salterio, S., and A. Webb, 2004. "The Balanced Scorecard: The Effects of Assurance and Process Accountability on Managerial Judgment." *The Accounting Review*, 1075-1094.

Lipe, M. G., and S. Salterio, 2000. "The Balanced Scorecard: Judgmental Effects of Common and Unique Performance Measures." *The Accounting Review*, 283–98.

Martinson, M., R. Davison, R., and D. Tse, 1999. "The Balanced Scorecard: A Foundation for the Strategic Management of Information Systems." *Decision Support Systems,* 71-88.

McClave, J., P. Benson, P., and T. Sinich, 2000. *A First Course in Business Statistics 8th Edition*. New Jersey: Prentice Hall.

Nahartyo, E., and I. Utami, 2016. *Panduan Praktis Riset Eksperimen*. Jakarta: Indeks.

Neumann, B., M. Roberts, and E. Cauvin, 2010. "Information Search Using the Balanced Scorecard: What Matters?" *The Journal of Corporate Accounting & Finance*: 61-66.

Niven, P., 2002. *Balanced Scorecard Step by Step: Maximizing Performance and Maintaining Results*. New York: John Wiley & Sons, Inc.

O'Reilly, C., 1980. "Individuals and Information Overload in Organizations: Is More Necessarily Better?" *Academy of Management Journal*, 684-696.

Rigby, D., and D. Bilodeau, "Management Tools and Trends 2013." Bain & Company survey, 2013. Website, Accessed December, 18, 2015 from http://www.bain.com/Images/BAIN_BRIEF_Management_Tools_2015.pdf

Roberts, M., T. Albright, and A. Hibbets, 2004. "Debiasing Balanced Scorecard Evaluations." *Behavioral Research in Accounting*, 75-88.

Schaubroeck, J., and K. Muralidhar, 1991. "A Meta-Analysis of the Relative Effects of Tabular and Graphic Display Formats on Decision Making Performance." *Human Performance*, 127-145.

Swain, M., and S. Haka, 2000. "Effects of Information Load on Capital Budgeting Decisions." *Behavioral Research in Accounting*, 171-198.

Tversky, A., and D. Kahneman, 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science*, 1124-1131.

Umanath, N., and I. Vessey, 1994. "Multiattribute Date Presentation and Human Judgment: A Cognitive Fit Perspective." *Decision Science,* 795-824.

Utami, A., 2011. Pengaruh Keterlibatan Emosi Positif Berupa *Affective Trust* Dalam mengurangi Bias pengukuran umumPada Evaluasi Kinerja *Balanced Scorecards.* Tesis Fakultas Ekonomika dan Bisnis Universitas Gadjah Mada Yogyakarta. 1-131.

Vessey, I., 1991. "Cognitive Fit: A Theory-Based Analysis of the Graphs Versus Tables Literature." *Decision Sciences*, 219-240.