CHAPTER 3

RESEARCH HYPOTHESES AND RESEARCH DESIGNS

3.1 Experiment I: The "Commitment Effect"

3.1.1 Hypothesis Development

The focus of this study is on whether involvement may influence performance evaluation of managers. The following hypothesis is based on the discussion in the previous section.

For the benchmark condition, I explore the situation when management does not have any involvement in the development process of the BSC (i.e. mute condition). Lipe and Salterio (2000) explore managers' performance evaluations for RadWear and WorkWear, the two women's apparel divisions for WCS Incorporated. They found that when common measures favor RadWear over WorkWear, participants evaluate RadWear's manager as having significantly better performance. Similarly, when common measures favor WorkWear over RadWear, participants evaluate WorkWear's manager as having significantly better performance. On the other hand, when unique measures favor RadWear or WorkWear, there is no significant difference in the evaluations of managers of the two divisions. Thus, Lipe and Salterio's (2000) result shows that there is common measure bias; i.e. that unique measures of a business unit are underweighted in the performance evaluation of managers.

As mentioned in the previous section, there are different degrees of involvement: (1) voice, (2) choice, and (3) voice and choice (Hunton and Price, 1994). For the purpose of this experiment, I choose to explore only the voice condition, which is a condition whereby a person has probabilistic control over the decision-making process, since the extent that voice impacts decision outcome is uncertain. This is appropriate in the context of this study which explores performance evaluations using the BSC. In organizational settings, the provision of choice does not guarantee that everyone receives the chosen choices (Baldwin et al. 1991). The same holds true for the development process of the BSC. Top management is not the

one who comes up with the BSC measures, but they are rather the ones who approve and agree to the measures used for each business unit. Divisional managers are the ones who come up with the BSC measures that are unique to their divisions (Lipe and Salterio, 2000). Also, not every top management agrees to every BSC measure used in the firm and divisions. The BSC measures developed by each division are rather a consensus from the top management.

As stated in the literature review, prior research in psychology found that involvement leads to attitude formation or attitude change (Park et. al., 2007), which then leads to commitment to a specific issue (Crosby and Taylor, 1983; Iwasaki and Havitz, 1998). Commitment usually results after a person is involved with an issue. This is due to a "fair process effect", which states that a person is more likely to accept decisions and their consequences if they are involved in making them (Folger et al., 1979). Many accounting researchers also explore the concept of involvement in their research. Tan (1995) explores how audit decision process is affected by prior expectations, prior audit involvement, and the review process. Results suggest that auditors with prior audit involvement paid more attention to consistent fact than to inconsistent facts, compared to auditors who just took over the audit of the firm from previous auditors. According to Tan (1995), involvement causes auditors to recall consistent facts than inconsistent facts. Elliott et al. (2004) propose that individuals complete seven steps when conducting financial analysis. The steps include defining the objective of the analysis, defining relevant information, acquiring information, evaluating and combining information with prior knowledge, using knowledge to make judgments, making a decision based on judgments, and carrying out the decision. If managers are involved with the development of the BSC, they will be automatically forced to exercise these seven steps before they actually evaluate their subordinates' performance. Thus, they will be more likely to incorporate both common and unique measures in their performance evaluations, since they are more acquainted with the BSC and they understand the measures more thoroughly from exercising the seven steps before making performance evaluations. Libby and Luft (1993) suggest that participants' performance on an experimental task depends on participants experience, ability, and knowledge.

Based on the above explanations, I hypothesize that when managers are involved in the development of the BSC, they are more likely to accept decisions and their consequences, even though the BSC measures may not reflect the choices that they make. In other words, involvement in the development of the BSC leads to commitment to the usage of both common and unique measures of the BSC. Also, involvement allows managers to acquire and integrate knowledge and experience about the BSC measures better than managers who are not involved in the development process. Based on the above discussion, I propose the following hypothesis:

H1: When managers are involved in the development process of the BSC, performance evaluations using the BSC will be affected by both common and unique measures.

3.1.2 Experimental Task

I use the experimental case developed by Lipe and Salterio (2000) with modifications. Participants were to assume a role of top management at WCS Incorporated, a firm specializing in women's apparel. The case describes that WCS has recently implemented the BSC and describes key features of the BSC. The case involves WCS's two largest divisions: RadWear, a retail division that focuses on clothing for urban teenagers, and Workwear, a division that sells business uniforms directly to clients. The BSC measures include measures that are common to the two divisions and measures that are unique to each of the two divisions. There are all together 16 measures, separating into four measures for each of the four perspectives: learning and growth, internal business processes, customer, and financial. For each perspective, there are two common measures and two unique measures. After reading the case, participants are to separately evaluate the performance of RadWear's manager and WorkWear's manager on a scale of 0-100, with 100 being excellent performance. Then, participants are required to compare the performance of both managers and recommend one manager to promote as manager of sales operation.

Experimental Design

The experiment is a 2 x 1 between-subjects design. The between-subjects factor is whether managers are involved in the development of the BSC (*involvement*, *no-involvement*). Consistent with Lipe and Salterio (2000), managers have to evaluate the performance of two divisions: RadWear and WorkWear, using two different sets of BSC. Lipe and Salterio (2000) explore four settings. In the first setting, both common and unique measures favor RadWear. In the second setting, both common and unique measures favor WorkWear. In the third setting, common measures favor RadWear, while unique measures favor WorkWear. In the fourth setting, common measures favor WorkWear, while unique measures favor RadWear. Similar to Libby et al. (2004), for this experiment I choose not to explore the first two settings, whereby both common and unique measures favor either one division. This is because the results found by Lipe and Salterio (2000) for the two settings are unequivocal. Thus, following Libby et al. (2004), I explore only the setting where performance on common measures favors RadWear and performance on unique measures favors WorkWear. Consistent with Lipe and Salterio (2000), common and unique measures had the same excess performance. The sum of excess performance for common measures is approximately 85 for RadWear and 52 for WorkWear (a difference of about 33). The sum of excess performance for unique measures is 85 for WorkWear and 52 for RadWear (a difference of about 33). So, if participants rely more on the common measures, they will evaluate RadWear's manager as having better performance than WorkWear's (i.e. RadWear evaluation minus WorkWear evaluation will be significantly greater than zero), and vice versa. Consistent with Libby et al. (2004), we do not vary the order of division presentation or manager evaluation, since Lipe and Salterio (2000) found that there is no order effect. Examples of the full case material for Experiment I are found in the Appendix A.

Experimental Conditions

As discussed above, the independent factor is participant's involvement in the development of the BSC measures for the business units. Involvement is manipulated at two levels: participants are involved and are not involved in the development of the BSC measures. Lipe and Salterio (2000) specify in the WCS case that "the top management were to meet with each divisional manager to communicate the firm-wide mission and to discuss with each manager that it is the division manager's role in developing a BSC for his own division. After the divisional scorecards were developed, each divisional manager met again with top management to explain his division's scorecard, to answer questions, and to make necessary adjustments as requested by the top management." However, Lipe and Salterio (2000) did not actually implement these steps in their research. Instead, they developed the BSC measures and let participants use the measures to evaluate subordinates. As hypothesized above, these steps of involvement by top management are crucial to the successful use of the BSC in the firm. Thus, I use the above steps outlined by Lipe and Salterio (2000) in the experiment. In this experiment, participants are to take the role of top management team of WCS. They have to help the CFO choose the appropriate unique measures for RadWear and WorkWear divisions.

Participants are randomly divided into two groups: one having involvement and the other having no involvement in the development process of the BSC. For the group that has involvement, participants are provided with the case about WCS, which describes the firm's background, mission, each division's details, and common measures that are used by both divisions. The case explains that some measures are similar for WorkWear and RadWear. These measures are already decided upon by the top management. The case further explains that since the two divisions are quite different in terms of sales methods and performance targets, some of the measures are different for the two divisions. Participants are asked to help the CFO decide upon the measures that are used differently for the two divisions (i.e. unique measures) by rating the appropriateness of the measures for RadWear and WorkWear. Participants are provided with a list of five unique measures for RadWear and five unique measures for WorkWear. Participants are to rate the unique

measures by assigning a score of 1, 2, 3, 4, or 5 to each of the measures, with 1 being the least appropriate unique measure and 5 being the most appropriate unique measure for each of the perspective. They also have to write a reason under each perspective as to how they rank the measures. This allows them to think thoroughly before actually rating the measures. The case did not specifically tell participants that the "common measures" are given and they have to rate the "unique measures". This is because I do not want participants to be able to guess what I am manipulating in this experiment. Telling participants too specifically about "common" and "unique" measures is too obvious and may confound the results if participants are able to guess the manipulation of "common" and "unique" measures.

Regression Models and Variables Definitions

To test H1, I estimate the following regression model:

$$DIFSCORE = \alpha + \beta_1 CME + \varepsilon_i \qquad (1)$$

where:

DIFSCORE = the difference between participants' evaluations of the two divisions (i.e. RadWear – WorkWear)

CME = "Commitment effect," which is a variable representing the mean differences between the *involvement* versus the *no-involvement* conditions

DIFSCORE is the difference between participants' evaluations of the two divisions. Participants are to evaluate the performance of manager of RadWear and manager of WorkWear on a scale of 0 – 100, with 0 being reassign and 100 being excellent performance. Then, participants are required to compare the performance of both managers and recommend one manager to promote as manager of sales operation. Consistent with Libby et al. (2004), the differences in participants' evaluations of the two divisions serve as the dependent variable. Since RadWear (WorkWear) outperforms WorkWear (RadWear) on common (unique) measures and if participants rely more on the common measures, a positive difference between

RadWear evaluation and WorkWear evaluation is indicative of that reliance, since the total percentage above target across all measures are the same for each division. If involvement can increase the use of unique measures in the performance evaluation, the differences in participants' evaluations should be reduced.

For *CME*, which is the independent variable, I contrast code the *no-involvement* and the *involvement* conditions as 1 and -1, respectively. Cohen (1968) discusses the use of contrast coding in linear regression models for use in the analysis of variance (ANOVA) and analysis of covariance (ANCOVA). Cohen (1968) stated that the main affects and interaction of ANOVA and ANCOVA can be reflected in a linear model through the use of specifically coded predictor vectors. Lewis and Mouw (1972) stated that "the contrast coding is a system that provides a logical and relatively simple method for developing regression models to answer more specific questions than the overall main effects and interaction tests generally applied in ANOVA". Actually for this simple regression analysis, I can use a dummy variable (i.e. 0 and 1) to represent the *no-involvement* and the *involvement* conditions. However, I use contrast coding, since I also use contrast coding for Experiment II. Using contrast coding in both Experiment I and Experiment II will make my analysis more consistent between the two experiments. The contrast coding assigns the two conditions as 1 and -1, since these are the standard coefficients for orthogonal polynomials; i.e. the sum of the codes assigned must be zero. If the coefficient of *CME* is significant, it will support H1 that involvement leads to commitment, which eventually causes participants to rely on both common and unique measures when they evaluate their subordinates using the BSC.

Control Variable: Knowledge (KNOW)

Prior knowledge may affect judgment of a decision maker. Prior knowledge is a major intervening variable that may affect judgment of a decision maker (Locke and Schweier, 1979; Park et al., 2007; and Iwasaki and Havitz, 1998). Kennedy (1995) explores the impact that the "curse of knowledge" has on audit judgment. "Curse of knowledge" occurs when decision makers are not able to ignore information they already processed. The more knowledge the person has about a particular subject matter, the more unnatural he becomes when making judgment

about the subject matter. Kennedy (1995) explores the effect that "curse of knowledge" may have on the judgment of auditors about going concern evaluation and analytical review. Results show curse of knowledge exists and it is not mitigated by accountability.

In the BSC context, prior knowledge that each participant has about the BSC concept and application may affect performance evaluation of participants using the BSC. Participants may develop this prior knowledge about BSC from many sources, such as work experience, academic study (Libby et al. 2004), readings in magazine articles and books, seminars on BSC, and etc. Thus, the best way to represent this knowledge is to directly test how much participants know about the BSC concept. So, I use the knowledge test to do this job. The knowledge test comprises eight multiple choice questions, which test the BSC concept and applications. The knowledge test asks questions about the underlying concept of the BSC, linkages of each of the four BSC perspectives, strategic execution of the BSC, steps in implementing the BSC, measures of the BSC, and performance evaluation using the BSC. Participants are required to perform the knowledge test prior to reading the WCS case. A knowledge test is found in Appendix A.

I use the following regression model controlling for prior knowledge about BSC of participants to estimate H1:

$$DIFSCORE = \alpha + \beta_1 CME + \beta_2 KNOW + \varepsilon_i \qquad (2)$$

where:

KNOW = mean of scores that participants received from the knowledge test about BSC concept and application, ranging from 0 – 8

Other variables are already defined in Equation (1).

Experimental Procedure

For the *involvement* condition, participants are given three parts of experimental instruments. Participants are required to complete Part I before proceeding to Part II and Part III. Part I is the knowledge test, which tests the general

knowledge of participants about the BSC concept. After completing the Part I, participants proceed to read Part II, which is the case about WCS Inc., describing the firm's background, mission, and each division's details. Participants have to rate the appropriateness of unique measures for each of the four BSC perspectives for RadWear and WorkWear using a score of 1 – 5. After completing this task, participants proceed to Part III, whereby I use the BSC that was developed by Lipe and Salterio (2000). Participants are given BSCs for RadWear and WorkWear. They have to evaluate the performance of RadWear and WorkWear managers using the given BSCs and recommend one manager for job promotion.

There are several reasons why I use Lipe and Salterio's (2000) case instrument. First, using the case instrument developed by Lipe and Salterio (2000) increases validity of the case material. As stated in the literature review, Lipe and Salterio (2000) were the first to explore performance evaluation using the BSC. Research that follows include Banker et al. (2004), Libby et al. (2004), Roberts et al. (2004), and Dilla and Steinbart (2005). Libby et al. (2004), Roberts et al. (2004), and Dilla and Steinbart (2005) use the case material developed by Lipe and Salterio (2000), while Banker et al. (2004) adopted Lipe and Salterio (2000) case material with adjustments. Second, the purpose of the involvement condition is to allow participants to be involved in the development process of the BSC, in order to create commitment to the usage of BSC measures in performance evaluations. Also, Baldwin et al. (1991) states that not everyone receives their choice in the organizational settings. Thus, whether the participants receive the unique measures that they choose should not matter, since the objective of the involvement condition is only to create a setting whereby participants are involved with the development process of the BSC. This method is realistic for involvement, since in the real firm setting, top management are not the ones who come up with the BSC measures, but they are the ones who approve and make comment on the unique measures used for each division. Divisional managers are the ones who come up with the BSC measures that are unique to their divisions (Lipe and Salterio, 2000). Also, not every top management agrees to every BSC measure used in the firm and divisions. The BSC measures developed by each division are rather a consensus from the top

management. Participants are told that they may be contacted by the author about their rankings. This is to create accountability for the participants.

Participants in the *no involvement* condition are given only two parts of the experimental instrument. Participants are required to complete Part I before proceeding to Part II. Part I is the knowledge test, and Part II is the WCS case. Participants in this condition get the same 16 BSC measures as participants in the *involvement* condition. However, the difference is that they do not take part in rating the appropriate and preferred unique BSC measures. The participants' task is to evaluate the two managers from RadWear and WorkWear based on the BSC measures.

3.1.3 Participants

Sixty-three M.B.A. students from a leading public university in Thailand participated in this experiment. Participants were in their second year of study. Table 1 presents descriptive statistics about the experimental participants. Average age of participants is 27.5 and the average number of years of working experience is 5.2 years. Participants have a mean knowledge test score of 4.3 from the full score of 8 points. This means that participants have a medium knowledge of the BSC concept and application, which is sufficient for the performance of this experiment. Majority of participants (66.7%) in this experiment have taken a course that teaches about the BSC. Moreover, 9.5% of participants have a real experience using the BSC to evaluate their subordinates, and 23.8% of participants have been evaluated by their superiors using the BSC in the organizations that they work for.

TABLE 1
Descriptive Statistics of M.B.A. Students Participating in Experiment I

Number of participants:	63
involvement condition	33
no-involvement condition	30

	Mean (Standard Deviation)
Age (years)	27.5 (2.9)
Work experience (years)	5.2 (2.5)
Number of subordinates	2.2 (5.86)*
Gender (male percentage)	42.9

Work experience:	Number of participants	%
Accounting, auditing or taxation	8	12.7
Finance, banking or investing	10	15.9
Marketing or sales	11	17.5
General management or personnel	10	15.9
Engineering	13	20.6
Medical	0	0.0
Others	11	17.5
	63	100%

	Mean (Standard Deviation)
Knowledge test score (ranging from 0 - 8 points)	4.3 (1.5)

Experience of participants with BSC:	Number of participants	%
Have used BSC to evaluate subordinates	6	9.5
Have been evaluated by BSC	15	23.8
Have studied a course on BSC	42	66.7

* There is an outlier for this item. One participant is a business owner, who has 150 subordinates. Before getting rid of outier, the mean (standard deviation) for the number of subordinate is 4.5 (19.5).

3.2 Experiment II: the "Frustration Effect"

3.2.1 Hypotheses Development

As explained in the literature review, some researchers contend that involvement often leads to commitment to a specific issue or object (Crosby and Taylor, 1983; Iwasaki and Havitz, 1998). On the other hand, some researchers contend that involvement may lead to a feeling of frustration, which has a negative effect on commitment (Folger et al., 1979). Referring to Experiment I, there may still be some questions about involvement. Although participants are involved with the development process of the BSC, they do not receive all the choices of unique measures that they choose for the BSC of each business unit. This is because I use the BSC developed by Lipe and Salterio (2000), which may include measures that are not chosen by the participants. Thus, one may question about whether the degree of inclusiveness of choices may lead to different levels of satisfaction, which may lead to either "commitment" or "frustration effect". Baldwin et al. (1991) explores the effects of trainee choice of training on subsequent motivation and learning based on three conditions: (a) no choice of training; (b) choice of training, but choice not received; (c) choice of training with choice received. Results show that trainees who have a choice of training and were given a training of their choice have greater motivation to learn. However, trainees who have a choice of training but did not receive the training of their choice were even less motivated than trainees who were not provided with the training choice. Applying the result of this paper to performance evaluations using the BSC, this may imply that the different degree of inclusiveness of choices may lead to different degree of usage of unique measures.

For this experiment, I explore the conditions when all managers have involvement in the development process of the BSC, but they do not all receive the choices that they make for the BSC measures. The first condition is when managers are provided with a choice of the BSC and they receive the particular choice that they make. This condition is based on the "intrinsic motivation theory", which contends that choice is inherently a good thing, since it is crucial to one's feeling of mastery and self-determination (Hicks and Klimoski, 1987). Thus, when managers are provided with a choice of BSC and receive the particular choice that they choose, they

will feel as if they have control over the decision process, which leads to commitment to the chosen BSC. This is consistent with Salancik (1977) who found that increased commitment occurs under conditions of participation and choice.

The second condition is when managers are provided with a choice of the BSC, but they do not receive that choice. This condition is realistic, since in organizational settings, the provision of choice does not guarantee that everyone receives their choice (Baldwin et al., 1991). Also, prior literature found that involvement has its limitations; i.e. involvement does not always enhance satisfaction with decision outcomes. Folger et al. (1979) labeled this as the "frustration effect". "Frustration effect" raises one's expectations, in which if the choice is not accepted, the person may be dissatisfied with the outcome. In the BSC context, if managers are provided with a choice of BSC, but their choice is not accepted, managers may be dissatisfied, which may have negative impact on performance evaluations using the BSC. They may be less motivated to evaluate subordinates based on the given BSC. Thus, their evaluation will be more affected by the common measures than by the unique measures, since common measures are easier to use under comparative evaluations (Slovic and MacPhillamy, 1974).

The third condition is when managers are provided with a choice of the BSC, but they only receive half the choices that they choose. As explained above, this condition is realistic, since in organizational settings, the provision of choice does not guarantee that everyone receives the choice that they make (Baldwin et al., 1991). In this case, it is difficult to predict the results of performance evaluations of managers using the BSC. Salancik (1977) suggests that increased commitment occurs when there is participation and choice. However, Folger et al. (1979) and Baldwin et al. (1991) suggest that there may be perils of participation; i.e. a person may be frustrated when his choice is not accepted. This means that involvement may not always lead to commitment. So, if managers are provided with a choice of the BSC, but receive only some of their chosen choices, they may be satisfied or dissatisfied with the outcome. If managers are satisfied with the outcome, they will be more committed and motivated to use the BSC in the performance evaluations. Thus, they will tend to use both common and unique measures in the performance evaluations. However, if managers are dissatisfied with the outcome, they may be frustrated and

ignore unique measures in the performance evaluations. I hypothesize that participants will base their performance evaluations more on common measures, since common measures are comparable across the two divisions and easier to use than unique measures (Lipe and Salterio, 2000). However, managers may be affected by the different degrees of choice received; i.e. they may be more frustrated when they do not receive all their chosen choices, comparing to when they receive some of their chosen choices. If managers are very frustrated with not receiving their choices at all, they may not believe in the BSC measures. This may lead them to rely more on common measures in their performance evaluations, since common measures are easier to use in a comparative situation. Based on the above discussion, I propose the following two hypotheses:

H2: When managers are provided with a choice of BSC but do not receive all the choices that they choose (i.e. receive half or none), they will base their performance evaluations more on common measures, comparing to when they are provided and receive all the choice that they choose.

H3: When managers are provided with a choice of BSC but do not receive the choice that they choose at all (i.e. receive none), they will base their performance evaluations more on common measures, comparing to when they are provided and receive some of the choices that they choose.

3.2.2 Experimental Task

Similar to Experiment I, I use the experimental case developed by Lipe and Salterio (2000) with modifications. Participants were to assume a role of top management at WCS Incorporated, a firm specializing in women's apparel. The case describes that WCS has recently implemented the BSC and describes key features of the BSC. The case involves WCS's two largest divisions: RadWear, a retail division that focuses on clothing for urban teenagers, and Workwear, a division that sells business uniforms directly to clients. The BSC measures include measures that are common to the two divisions and measures that are unique to each of the two divisions. There are all together 16 measures, separating into four measures for each

of the four perspectives: learning and growth, internal business processes, customer, and financial. For each perspective, there are two common measures and two unique measures. After reading the case, participants are to separately evaluate the performance of RadWear's manager and WorkWear's manager on a scale of 0-100, with 100 being excellent performance. Then, participants are required to compare the performance of both managers and recommend one manager to promote as manager of sales operation.

Experimental Design

The experiment is a 3 x 1 between-subjects design. The between-subjects factor compares the different degrees of managers' choice received. More specifically, every participant is involved in the development process of the BSC. However, the choices that they choose for the BSC during the development process may not be reflected in the final version of the BSC that they have to use to evaluate their subordinates. The degree of choice received is manipulated at three levels: (1) all choices are received, (2) half the choices are received, and (3) none of the choices is received. Similar to Experiment I, I examine the setting where performance on common measures favors one division and performance on unique measures favors the other division. An example of the full case material for Experiment II is found in the Appendix B.

Experimental Conditions

As discussed above, the independent factor is the different degrees of participant's choice received. Unlike Experiment I, all participants in this experiment are involved in the development process of the BSC. Participants are to take the role of top management team of WCS. They have to help the CFO choose the appropriate unique measures for RadWear and WorkWear. Participants are provided with the case about WCS, which describes the firm's background, mission, detail about RadWear and WorkWear, and common measures that are similarly used for the two divisions. Similar to Lipe and Salterio (2000), each of the four perspectives contained four measures, with two being common and two being unique measures. The common measures are the same for both divisions and the same for the two sets.

Unlike Experiment I, participants do not have to rank and choose their preferred unique measures for each of the four BSC perspectives. Instead, participants are provided with two sets of unique measures for each business division. Each set contains two unique measures for each of the four perspectives. Unique measures are different for each division and are different for the two sets. I choose to manipulate the measures this way, since it is impossible to individually let participants rank their preferred unique measures and be able to explore the different levels of their choice received. Similar to Experiment I, the case did not specifically tell participants that some of the measures are "common measures" and that participants have to choose "unique measures". This is because I do not want participants to be able to guess what I am manipulating in this experiment. Telling participants too specifically about "common" and "unique" measures is too obvious and may confound the results.

Participants are randomly divided into three groups. The first group *(choice-all-received)* is the condition when participants receive the set of BSC that they choose. The second group *(choice-not-received)* is the condition when participants do not receive the set of BSC that they choose. The third group *(choice-half-received)* is the condition when participants receive half of the BSC measures that they choose. Participants are provided with two sets of BSCs to choose from; i.e. set A and set B. They have to write a reason why they choose the given set. This allows them to think thoroughly before actually choosing between the two sets. Participants are also told that they may be contacted by the author about their choice. This is to create accountability for the participants, so that they may be more careful with performing the experiment.

For the *choice-all-received* condition, participants receive the set of BSC that they choose. For example, if a participant chooses set A, he will be given set A of BSC as measures to evaluate managers from RadWear and WorkWear. On the other hand, if a participant chooses set B, he will be given set B of BSC as measures to evaluate managers from both divisions. Participants are told that they receive the choice that they choose for the BSC and that the final version of BSC provided for evaluation is a consensus from the top management team of WCS.

For the *choice-not-received* condition, participants do not receive the set of BSC that they choose. For example, if a participant chooses set A, he will be given

set B of BSC as measures to evaluate and compare managers from the two divisions. On the other hand, if a participant chooses set B, he will be given set A of BSC to evaluate and compare managers from both divisions. Participants are told that they do not receive the choice that they make for the BSC and that the given set of BSC provided for evaluation (i.e. a set that participant does not choose) is a consensus from top management team of WCS.

For the *choice-half-received* condition, participants receive half of what they choose. If participants choose set A or set B, they will be given one unique measure from set A (i.e. measure that they choose) and one unique measure from set B (i.e. measure that they did not choose) for each of the four perspectives. This is to avoid confounding effect from participants receiving only the unique measures that they choose for one perspective, while not receiving the unique measures that they choose for the other perspective. Participants are told that they receive half of what they choose, since the given set of BSC provided for evaluation is a consensus from the top management team of WCS.

Regression Model and Variable Definitions

Based on the above discussion, I estimate the following regression model to test H2 and H3:

$$DIFSCORE = \alpha + \beta_1 INE + \beta_2 FFE + \beta_2 MFE + \varepsilon_i \qquad (3)$$

where:

DIFSCORE	=	the difference between participants' evaluations of the two divisions (i.e. RadWear – WorkWear)
INE	=	"involvement effect," which is a variable comparing mean differences between *no-involvement* condition with *choice-all-received*, *choice-half-received*, and *choice-not-received* conditions
FFE	=	"full frustration effect," which is a variable comparing mean differences between *choice-all-received* condition with *choice-not-received* and *choice-half-received* conditions

MFE = "marginal frustration effect," which is a variable comparing mean differences between *choice-half-received* condition with *choice-not-received* condition

DIFSCORE is the difference between participants' evaluations of the two divisions. Participants are to evaluate the performance of manager of RadWear and manager of WorkWear on a scale of 0 – 100, with 0 being reassign and 100 being excellent performance. Consistent with Libby et al. (2004) and Experiment I, the difference in participants' evaluations of the two divisions serve as the dependent variable. Since RadWear (WorkWear) outperforms WorkWear (RadWear) on common (unique) measures and if participants rely more on the common measures, a positive difference between RadWear evaluation and WorkWear evaluation is indicative of that reliance, since the total percentage above target across all measures are the same for each division. So, if different levels of choice received affect the usage of unique measures in performance evaluation, the differences in participants' evaluations of the two managers should be reduced.

To test H2 and H3, I contrast code the three experimental conditions: *choice-all-received*, *choice-half-received*, and *choice-not-received* and the *no-involvement* condition in Experiment I into three variables as follows[1]:

	no-involvement condition	*choice-not-received condition*	*choice-half-received condition*	*choice-all-received condition*
INE	-3	1	1	1
FFE	0	1	1	-2
MFE	0	-1	1	0

INE ("involvement effect") compares the mean differences of the *no-involvement* condition with the other three conditions, in which all participants have equal involvement in the development process of the BSC. I predict that the mean

[1] Contrast coding is one way of coding, which yields the same statistical inferences as coding with dummy variables.

difference of the *no-involvement* condition should be larger than the mean differences between the *choice-not-received*, *choice-half-receive*d, and *choice-all-received* conditions. This is because I predict that the *no-involvement* participants should base their performance evaluations on the common measures, following Lipe and Salterio (2000), whereas participants in the other three conditions should base their performance evaluations on both common and unique measures, since involvement in the development of the BSC measures leads to commitment to both common and unique BSC measures. So, the predicted sign for the coefficient of *INE* is negative. If the coefficient of *INE* is statistically significant, it means that the difference in performance evaluations of participants in the *no-involvement* condition is different from the evaluations of the other three conditions. If this is true, the result will support H1 that involvement leads to commitment, which causes participants to use both common and unique measures in their performance evaluations using the BSC.

FFE ("full frustration effect") compares the mean differences of the *choice-all-received* condition with *choice-not-received* and *choice-half-received* conditions. In other words, *FFE* compares the condition in which participants receive all the choices of BSC measures that they choose with the conditions in which participants voice their opinion about the BSC measures, but receive only some or none of the BSC measures chosen. I predict that the coefficient of FFE should be positive, since participants in the *choice-all-received* conditions are more likely to use more of the unique measures in performance evaluations, compared to participants in the other two conditions, who receive only some or none of the choice received. Thus, this will bring the mean difference between RadWear and WorkWear of the *choice-all-received* condition closer to zero. On the other hand, when participants receive some or none of their chosen choices, they may be frustrated with their choices not being received. Thus, they may be more likely to base their performance evaluations more on the common measures. So, the mean differences between RadWear and WorkWear for the *choice-not-received* and *choice-half-received* conditions will be larger than that of the *choice-all-received* condition. If the coefficient of *FFE* is statistically significant, it means that the difference in performance evaluations of participants in the *choice-all-received* condition is

different from the evaluations of the other two conditions. If this is true, this will support H2 that "frustration effect" occurring from not receiving all the choices that participants choose causes them to rely more on the common measures when they evaluate their subordinates.

MFE ("marginal frustration effect") compares the *choice-not-received* condition with the *choice-half-received* condition. Whether managers receive half or none of the choices that they choose, they are affected by the frustration effect. However, the degree of this effect may be different. That is, when managers do not receive the choice that they choose at all, they may be more frustrated than when they receive half of what they choose. So, managers who receive none of their chosen choices may rely more on the common measures, since frustration may lead them to have stronger disbelief in the BSC measures, compare to when they receive some of their chosen choices. Base on this discussion, I predict that the coefficient of *MFE* should be negative, since participants in the *choice-not-received* condition may rely more on the common measures than those in the *choice-half-received* condition. If the coefficient of *MFE* is significant, this will support H3 that different degrees of choice not received lead to different levels of frustration.

Control Variable: Knowledge (KNOW)

As explained in Experiment I, prior knowledge may affect judgment of a decision maker (Kennedy 1993, 1995). In the BSC context, prior knowledge that each participant has about the BSC concept may affect performance evaluation of participants using the BSC. Participants may develop this prior knowledge about BSC from many sources, such as work experience, academic study (Libby et al. 2004), readings in magazine articles and books, seminars on BSC, and etc. Following Experiment I, I use knowledge test to measure how much participants know about the BSC concept. The knowledge test comprises eight multiple choice questions, which test the BSC concept and applications. The knowledge test asks questions about the underlying concept of the BSC, linkages of each of the four BSC perspectives, strategic execution of the BSC, steps in implementing the BSC, measures of the BSC, and performance evaluation using the BSC. Participants are required to perform the knowledge test prior to reading the WCS case.

Thus, I use the following regression model controlling for prior knowledge about BSC of participants to estimate H2 and H3:

$$DIFSCORE = \alpha + \beta_1 INE + \beta_2 FFE + \beta_2 MFE + \beta_2 KNOW + \varepsilon_t \qquad (4)$$

where:

KNOW = mean of scores that participants received from the knowledge test about BSC concept and application, ranging from 0 – 8

All other variables are as described in Equation (3).

Experimental Procedure

As stated above, all participants in Experiment II are equally involved in the development of the BSC. Participants are given three parts of the experimental instruments. However, for this experiment, participants are first provided only two parts; i.e. Part I and Part II. After completing Part II, they will be given Part III of the experimental instrument.

Participants are required to complete Part I before proceeding to Part II and Part III. All participants receive the same Part I and II of the case instrument, but differs in Part III of the case instrument. Part I is the knowledge test, which tests the general knowledge of participants about the BSC concept. After completing Part I, participants proceed to read Part II, which is the case about WCS Inc., describing the firm's background, mission, and each division's details. Participants are provided with the common measures and are told that the measures are already decided upon by the management of WCS. Participants are provided with two sets of unique measures (Set A and Set B) for RadWear and WorkWear. Participants have to choose between measures in Set A and Set B as to which set provides the measures that are most appropriate for the mission and strategies of each division. Participants are required to write down the reasons of their choice. This allows them to think thoroughly before deciding upon the set.

For Part III of the experimental instrument, participants in the *choice-all-received* condition will receive the particular set that they choose. For example, if

they choose Set A, I will give them Set A of the experimental instrument. Set A is the exact same case instrument as was used by Lipe and Salterio (2000). Set B includes the same common measures as those in Lipe and Salterio (2000), but differs from Lipe and Salterio (2000) in the unique measures. The unique measures found in Set B are developed by me from referring to accounting textbooks about BSC and business articles. For the *choice-not-received* condition, participants will not receive the choice that they choose. For example, if they choose Set A, I will give them Set B of the case instrument. For the *choice-half-received* condition, participants will receive only half of what they choose. After completing Part III, participants have to answer general and demographic questions.

3.2.3 Participants

One hundred M.B.A. students from a leading public university in Thailand participated in this experiment. Participants were in their second year of study. Participants are on average 27.1 years old with 4.9 years of working experience. Table 2 presents descriptive statistics about the experimental participants. Participants have a mean knowledge test score of 5.1 from the full score of 8. This means that participants have a medium level of knowledge about BSC concepts and applications, which is sufficient for performing this experiment. 81% of participants in this experiment have studied a course that teaches about the BSC. Moreover, 10% of participants have used BSC to evaluate subordinates, and 20% of participants have been evaluated by their superiors using the BSC in the organizations that they work for.

TABLE 2
Descriptive Statistics of M.B.A. Students Participating in Experiment II

Number of participants	100
choice-all-received condition	36
choice-half-received condition	33
choice-not-received condition	31

	Mean (Standard Deviation)
Age (years)	27.1 (3.9)
Work experience (years)	4.9 (1.9)
Number of subordinates	2.0 (3.2)
Gender (male percentage)	35

Work experience:	**Number of participants**	**%**
Accounting, auditing or taxation	13	13
Finance, banking or investing	19	19
Marketing or sales	25	25
General management or personnel	14	14
Engineering	14	14
Medical	3	3
Others	12	12
	100	100%

	Mean (Standard Deviation)
Knowledge test score (ranging from 0 - 8 points)	5.1 (1.6)

Experience of participants with BSC:	**Number of participants**	**%**
Have used BSC to evaluate subordinates	10	10
Have been evaluated by BSC	20	20
Have studied a course on BSC	81	81