CHAPTER 4

FINDINGS AND RESULTS

4.1 Manipulation Checks for Experiment I and Experiment II

Manipulation checks show that participants think that the case is realistic and easy to understand (p-*value* < 0.01). Participants recognize that the two divisions employ different performance measures, sell to different markets, and should employ different performance measures (p-*value* < 0.01). Participants also think that the measures presented in the BSC are appropriate for the two divisions (p-*value* < 0.01). There is no difference across experimental conditions in the ease of understanding, case difficulty, and case realism (p-*value* > 0.50). These results are consistent with Lipe and Salterio (2000) and Libby et al. (2004).

Table 3 shows participants' mean performance evaluations and differences in performance evaluations of RadWear and WorkWear for all five experimental conditions. For the *no-involvement* condition, the mean performance evaluations for RadWear and WorkWear are 77.33 and 70.67, respectively. So, the difference in performance evaluations for the *no-involvement* condition is 6.66. To examine whether participants rely more on common measures than unique measures in their performance evaluations, I test whether the difference in participants' performance evaluations of RadWear and WorkWear is significantly greater from zero. Result in Table 3 shows that the difference of 6.66 is statistically different from zero (p-*value* < 0.01). This result is consistent with the result found in Lipe and Salterio (2000) that common measures affect participants' evaluations more than unique measures, given the condition that participants do not have involvement in the development process of the BSC.

TABLE 3
Participants' Mean Performance Evaluations and Differences

Mean Performance Evaluations (Standard Deviations) for RadWear and WorkWear's Managers by Experimental Conditions

	No-Involvement	Involvement	Choice-Not-Received	Choice-Half-Received	Choice-All-Received
RadWear	77.33 (7.96)	72.42 (12.76)	71.45 (7.98)	72.88 (9.52)	70.14 (11.18)
WorkWear	70.67 (10.40)	71.82 (13.28)	67.74 (10.15)	70.91 (12.21)	70.69 (12.20)
Difference: RadWear - WorkWear	6.66	0.60	3.71	1.97	-0.55
t-statistic and p-value associated with significance of difference from "0" test	t = 4.130; p = 0.000 n = 30	t = 0.299; p = 0.767 n = 33	t = 2.219; p = 0.034 n = 31	t = 1.132; p = 0.266 n = 33	t = -0.310; p = 0.758 n = 36

Table 3 shows magnitudes of the mean differences in performance evaluations of RadWear and WorkWear for the three experimental conditions in Experiment II. The magnitude of the differences in performance evaluations when participants receive all their chosen choices (*choice-all-received* condition) suggests that participants use both common and unique measures in their performance evaluations, since the mean difference is -0.55 and is not statistically different from zero (p-*value* = 0.758). When participants receive some of their chosen choices (*choice-half-received* condition), they also tend to rely on both common and unique measures in their performance evaluations. This is because the mean difference of 1.97 is not statistically different from zero (p-*value* = 0.266). However, the magnitude of the difference in performance evaluations when participants receive none of their chosen choices (*choice-not-received* condition) suggests that participants tend to rely more on common measures than on unique measures, since the mean difference is 3.71 and is statistically different from zero (p-*value* < 0.05). Further analysis of the independent sample t-test suggests that the mean difference in the *choice-not-received* condition is not statistically different from the *no-involvement* condition (p-*value* = 0.209), which gives additional support that when participants are given a choice but do not receive their chosen choices, they tend to rely on common measures in their performance evaluation, as much as they do when they are not given a choice at all. This is because when participants do not receive their chosen choices, they are frustrated with not receiving those choices. As a result, they tend to rely more on common measures, since unique measures are not what they choose. Besides, common measures are easier to use under comparative situation (Lipe and Salterio, 2000).

4.2 Results for Experiment I

Test of H1: the "Commitment Effect"

H1 hypothesizes that when managers are involved in the development process of the BSC, managers will use both common and unique measures in their performance evaluations using the BSC. I use simple regression to test this hypothesis. Table 4 reports the result of this regression analysis. The coefficient of *CME* is statistically significant (p-*value* < 0.05) and is positive as was predicted. Since I contrast code the differences in performance evaluations of participants in the *involvement* and the *no-involvement* conditions with -1 and 1, respectively, a significant coefficient of *CME* means that there is a significant difference between the performance evaluations in the *involvement* versus the *no-involvement* conditions. More specifically, managers who does not have any involvement in the development of the BSC tend to base their performance evaluations on common measures, as found by Lipe and Salterio (2000). However, managers who are involved in the development process of the BSC tend to base their performance evaluations on both common and unique measures. This can also be seen from the mean performance evaluations of the two conditions. Table 3 shows that the mean evaluation of the *no-involvement* group is 6.66, which is statistically different from zero (p-*value* = 0.000), whereas the mean evaluation of the *involvement* group is 0.60, which is not statistically different from zero (p-*value* = 0.767). Thus, this result supports H1 that when managers are involved in the development process of the BSC, they will use both common and unique measures in their performance evaluations using the BSC. Involvement in the development of the BSC measures leads to attitude formation, if a person does not have prior knowledge about the BSC, or attitude change, if a person already has prior knowledge about the BSC (Park et al., 2007). The attitude formation or attitude change causes commitment to a particular issue, since a person is more likely to accept decisions and its consequences if he had participated in making them (Folger et al., 1979). As a result, participants who are involved in the development process of the BSC are committed to using the BSC measures in evaluating their subordinates.

TABLE 4
Results of Testing H1: The Impact of "Commitment Effect" on Performance Evaluations of Mangers

$$DIFSCORE = \alpha + \beta_i CME + \varepsilon_i$$

Variables	Predicted Signs	Coefficients	t statistics	p-*value*
α		3.636	2.771	0.007
CME [a]	+	3.030	2.309	0.012**
N	63			
Adj. R Square%	0.065			

Variables are defined as follows:
Independent variable: *CME* = Variable representing mean differences between *involvement* and *no-involvement* conditions, coded as -1 and 1 respectively
Dependent variable: *DIFSCORE* = The difference between participants' evaluations of the two divisions (RadWear - WorkWear)

[a] This p-*value* is one-tail, since the prediction of this variable is directional.
***, **, and * indicates statisticlly significant at p-*value* < .01, .05, and .10, respectively.

I also control for the effect of prior knowledge in the regression analysis, since prior psychology literature found that success of an involvement effort depends on many intervening variables. One of the most important one is the cognitive factor, particularly prior knowledge of managers (Breckler, 1984; Kyle and Mowen, 2005). Table 1 shows that the mean score for the knowledge test is 4.3 (standard deviation = 1.5) out of a total score of 8 points. Table 5 shows result of this regression analysis. Controlling for prior knowledge of participants, the coefficient of *CME* is still statistically significant (p-*value* < 0.05). However, the control variable *KNOW* is not significant. This means that prior knowledge does not have any effect on performance evaluations of manager using the BSC. This result is consistent with that

of Libby et al. (2004) who use "familiarity with the BSC" as one of the control variables and found that the variable is not statistically significant.

TABLE 5
Results of Testing H1: The Impact of "Commitment Effect" on Performance Evaluations of Mangers, Controlling for Prior Knowledge

$$DIFSCORE = \alpha + \beta_1 CME + \beta_2 KNOW + \varepsilon_i$$

Variables	Predicted Signs	Coefficients	t statistics	p-*value*
α		-1.812	-0.443	0.659
CME [a]	+	3.113	2.388	0.010***
KNOW	?	1.268	1.405	0.165
N	63			
Adj. R Square %	0.080			

Variables are defined as follows:
Independent variable: *CME* = Variable representing mean differences between *involvement* and *no-involvement* conditions, coded as -1 and 1 respectively
Control variable: *KNOW* = Score that participants received from the knowledge test about the BSC concept and application, ranging from 0 - 8.
Dependent variable: *DIFSCORE* = The difference between participants' evaluations of the two divisions (RadWear - WorkWear)

[a] This p-*value* is one-tail, since the prediction of this variable is directional.
***, **, and * indicates statisticlly significant at p-*value* < .01, .05, and .10, respectively.

4.3 Results for Experiment II

Test of H2 and H3: "Full Frustration Effect" versus "Marginal Frustration Effect"

H2 and H3 hypothesize about the impact of "frustration effect" on performance evaluations of managers using the BSC upon different levels of their choice received of the BSC measures. I use multiple regression to test these hypotheses. Table 6 reports the result of this regression analysis. The coefficient of *INE* is negative and statistically significant (one-tail p-*value* = 0.01) as expected. This means that performance evaluation of managers in the *no-involvement* condition is significantly different from that of managers in the *choice-all-received*, *choice-half-received*, and *choice-not-received* conditions. This result gives further support to H1 that when managers are involved in the development process of the BSC, they are more committed to the BSC, so they tend to use both common and unique measures in their performance evaluations of subordinates. Table 3 presents differences in participant's evaluations of RadWear and WorkWear. The mean differences for the *no-involvement* condition is 6.66, whereas the mean differences for the other three conditions is 1.60 (standard deviation = 10.12). When participants are not involved in the development process of the BSC, they tend to rely more on common measures upon their performance evaluations of subordinates. However, when they are involved in the development process of the BSC, regardless of whether they receive, does not receive, or receive only some of the choices that they choose, they seem to have commitment to the BSC of WCS. This is shown statistically by the significant difference (one-tail p-*value* < 0.01) between the *no-involvement* condition and the other three involvement conditions. Thus, "commitment effect" seems to be working for all involvement conditions, regardless of the degree of choice received.

The coefficient of *FFE* is positive and statistically significant (one-tail p-*value* = 0.05) as expected. This means that the difference in performance evaluations of participants in the *choice-all-received* condition is significantly different from that of the *choice-not-received* and the *choice-half-received* conditions. The result supports H2 that when participants are given a choice, but receive only half or none of the choice chosen, they will be frustrated and will tend to emphasize common

measures when they evaluate their subordinates. Table 3 presents differences in participant's evaluations of RadWear and WorkWear. The mean differences for the *choice-all-received* condition is -0.55, whereas the mean differences for the other two conditions is 2.81 (standard deviation = 9.63). The magnitudes of the mean differences suggest that performance evaluations of participants are affected by the different levels of choice received. When they receive all the choices that they choose, they feel as if they have control over the decision process, which leads to commitment to a chosen BSC. So, they tend to base their performance evaluations on both common and unique measures, since the mean difference of the *choice-all-received* condition is not significantly different from zero (given that 0 represents the condition when participants evaluate the performance of RadWear and WorkWear as indifferent), as shown in Table 3. This is consistent with the "fair process effect" that explains that a person is more likely to accept decisions and their consequences if he is involved in making them (Folger et al., 1979).

When participants do not receive the choices that they choose at all or receive only some of their chosen choices, results show a clear influence of the "full frustration effect", since the coefficient of *FFE* is positive and statistically significant (one-tail p-*value* = 0.05). Since participants in the *choice-all-received* condition use both common and unique measures in their performance evaluations of subordinates, a positive significant difference between *choice-all-received* condition and the other two conditions means that performance evaluations of managers in the *choice-not-received* and *choice-half-received* conditions are biased towards using common measures in their performance evaluations, which is as hypothesized in H2. This implies that when participants are given a choice, but does not receive all the choices that they choose or receive only some of the chosen choices, they become frustrated. This is because when a person is given a choice, he may expect that his choice should be accepted. So, when his choice is denied, he becomes dissatisfied with the subsequent outcome.

H3 is not supported, since the coefficient of *MFE* is not statistically significant, although the coefficient of *MFE* is negative as predicted. *MFE* compares the *choice-half-received* condition to the *choice-not-received* condition. So, when its coefficient is not significant, it means that both conditions are statistically equally

affected by the "frustration effect". I predict that the "frustration effect" should be stronger in the *choice-not-received* condition than in the *choice-half-received* condition, since participants may be more frustrated by not receiving all their chosen choices than receiving some of the chosen choices. However, the result does not support my prediction, which shows that the "frustration effect" in the *choice-not-received* condition is not as strong as expected. This may be explained by the fact that performance evaluation does not affect the well-being of participants, but affect subordinates who are being evaluated. More specifically, participants in this experiment do not face the same incentives as managers in the real business organizations, who are normally responsible for evaluating subordinates' performance. Managers in real business organizations have to report their performance evaluations to their bosses and are often questioned about their performance evaluations of subordinates. Besides, this experiment is conducted using a hypothetical situation, which has nothing to do with participants' personal lives or careers. So, even though participants do not receive the choices that they choose, the "frustration effect" is not as strong as when they are directly affected by not receiving the chosen choices.

The "frustration effect" found in this paper is not as strong as the "frustration effect" documented by Baldwin et al. (1991). Baldwin et al. (1991) offer opportunity for university trainees to enroll in a professional seminar that focuses on practical business skills. Trainees are randomly assigned to one of the three conditions: (1) no choice of training; (2) choice of training, but choice not received; and (3) choice of training with choice received. Result indicates that participants who are given a choice but were not given a training of their choice were significantly less motivated to learn than trainees who were not given a choice at all. The reason why "frustration effect" found by Balwin et al. (1991) seem to be stronger than that found in my paper is because participants in Baldwin et al. (1991) were real university trainees that do not receive the choice of the training program that they choose. So, their *choice-not-received* condition is real, applies to real participants (i.e. university trainees), and directly affects the participants' career and daily lives. Furthermore, participants have to actually participate in the training program that they are not interested. This clearly has a greater impact of the "frustration effect". Thus, there is

no doubt that under Baldwin et al. (1991) setting, the "frustration effect" seems to be stronger than that found in my paper.

TABLE 6
Results of Testing H2 and H3: The Impact of "Frustration Effect" on Performance Evaluations of Mangers

$$DIFSCORE = \alpha + \beta_1 INE + \beta_2 FFE + \beta_3 MFE + \varepsilon_i$$

Variables	Predicted Signs	Coefficients	t statistics	p-*value*
α		2.948	3.421	0.001
INE [a]	-	-1.240	-2.43	0.009***
FFE [a]	+	1.132	1.663	0.050**
MFE [a]	-	-0.870	-0.71	0.240
N	130			
Adj. R Square	0.047			

Variables are defined as follows:
Independent variables: *INE* = Variable which compares mean differences between *no-involvement* condition with *choice-all-received, choice-half-received* , and *choice-not-received* conditions
FFE = Variable which compares mean differences between *choice-received* condition with *choice-not-received and choice-half-received* conditions
MFE = Variable which compares mean differences between *choice-half-received* conditon with *choice-not-received* condition
Dependent variable: *DIFSCORE* = The difference between participants' evaluations of the two divisions (RadWear - WorkWear)

[a] This p-*value* is one-tail, since the prediction of this variable is directional.
***, **, and * indicates statisticlly significant at p-*value* < .01, .05, and .10, respectively.

Following Experiment I, I also control for the effect of prior knowledge in the regression analysis. Table 7 shows that, consistent with the results found in Experiment I, the coefficient of *KNOW*, which is the control variable, is not statistically significant. This means that prior knowledge does not have any effect on

performance evaluations managers using the BSC. This result is consistent with that of Libby et al. (2004), who use "familiarity of BSC" as one of the control variable and found the variable to be statistically insignificant.

TABLE 7
Results of Testing H2 and H3: The Impact of "Frustration Effect" on Performance Evaluations of Mangers, Controlling for Prior Knowledge

$$DIFSCORE = \alpha + \beta_1 INE + \beta_2 FFE + \beta_3 MFE + \beta_3 KNOW + \varepsilon_i$$

Variables	Predicted Signs	Coefficients	t statistics	p-*value*
α		2.752	0.945	0.346
INE [a]	-	-1.249	-2.367	0.010***
FFE [a]	+	1.141	1.640	0.052*
MFE [a]	-	-0.864	-0.701	0.242
KNOW	?	0.040	0.006	0.944
N	130			
Adj. R Square	0.040			

Variables are defined as follows:
Independent variables: *INE* = Variable which compares mean differences between *no-involvement* condition with *choice-all-received, choice-half-received* , and *choice-not-received* conditions
FFE = Variable which compares mean differences between *choice-received* condition with *choice-not-received and choice-half-received* conditions
MFE = Variable which compares mean differences between *choice-half-received* conditon with *choice-not-received* condition
Control variable: KNOW = Score that participants received from the knowledge test about the BSC concept and application, ranging from 0 - 8.
Dependent variable: *DIFSCORE* = The difference between participants' evaluations of the two divisions (RadWear - WorkWear)

[a] This p-*value* is one-tail, since the prediction of this variable is directional.
***, **, and * indicates statisticlly significant at p-*value* < .01, .05, and .10, respectively.