Developments in Business Simulation and Experiential Learning, Volume 30, 2003

THE IMPACT OF BUSINESS WAR GAMES: QUANTIFYING TRAINING EFFECTIVENESS

Christopher M. Scherpereel
Northern Arizona University
cms89@mail.cba.nau.edu

ABSTRACT

Does participation in a business simulation exercise change people's perspectives on decision problems? Does it change their approach to solving those problems? These are fundamental questions positively confirmed by this research. Using the semantic differential technique and a one-group pretest-posttest design, tests are conducted to assess changes in characterization and approaches to business decision problems. A matched pair t-test confirms significant change in ten of twelve key hypotheses; while the Wilcoxon signed-rank test confirms nine. The measurement methodology developed and results presented provide quantifiable justification for the use of business simulation exercises to induce targeted change in a decision maker's decision problem perception.

INTRODUCTION

"The average company spent about $10 million on internal and external executive development in 1998 … spending on U.S. corporate training and education for managers rose to $16.5 billion, up 17% from last year …" (Reingold, 1999).

Why do companies invest in training programs? In 1998, Archie W. Dunham, chairman and CEO of Conoco Inc., an oil and energy company, decided to make a major investment. With the price of oil and company revenues plummeting, Dunham decided to invest in his managers, rather than tighten spending and cut costs. "Even though oil was at $10, it was the right decision." How did Dunham justify this decision? Dunham's focus was on the future, "you're going to be successful long-term if you have good people" (Reingold, 1999). This intuitive justification is common in industry. At this extreme, Pete Peterson, vice president of personnel at Hewlett-Packard stated that "too much time, energy and creativity is spent on measuring training vs. accepting it based on face validity – and getting on with it" (Filipczak, Picard, & Stamps, 1998, p. 14). Unfortunately, substantial financial capital and time are being expended on training and there exists a need to provide some assessment of its value to the organization, even if it is only a qualitative measure of value. Thus, the question remains in the minds of most executives, what is the value (face value or otherwise) that I am getting from the money that I am spending?

The statement by Richard Farson sums up where the value is coming from, "education gives managers new ways of thinking, new perspectives … it can enable them to see the interconnectedness of events, to go beyond the conventional wisdom … to think strategically." (Farson, 1996, p. 156) This is equivalent to stating that management education is designed to get mangers to think differently and see things differently. Even with this insight, researchers and bean counters have continued to focus on developing instruments and designs to measure some tangible benefit. These "academic studies" focus on quantitatively measuring productivity changes and attempt to link any observed/measured changes to specific educational programs.

Although some success can be claimed in quantitatively measuring productivity gains from management development programs (Kirkpatrick 1994), the results are often suspect due to the great number of confounding parameters. David Fagiano (Fagiano, 1995), CEO of the American Management Association, suggests that quantitative measures should be limited to "Hard-skill courses such as 'Improving Your Word Processing Techniques' … (and) … technical courses such as 'System Analysis and Design'" (p. 12) where there are definable outcomes. The measurement problem becomes more difficult when the training has more subtle and longer-range payoffs, such as those associated with management development training. Therefore, more creative, alternative methodologies need to be explored.

Studies which focus on less quantitative measures of value, typically use questionnaires and interviews to elicit individual response to subjective questions. In these studies, the attempt is to measure training's value by measuring individual satisfaction. The assumption is that there exists a relationship between training satisfaction and job performance, and by measuring an individual's satisfaction with a training program the resultant job performance can be inferred. Studies supporting this relationship typically reference the established relationship between job satisfaction and job performance (Alliger et al., 1997; Iaffaldano & Muchinsky, 1985).

Individual's perceptions or attitude changes are sometimes measured as a proxy indicator of anticipated behavioral changes. Studies by Weigel, Vernon, & Tognacci

(1974) found that "Attitude measures should be expected to predict only behaviors that are appropriate to or specified by the attitude under consideration." (p. 728). This was confirmed by Ajzen and Fishbein (1977), whose results suggest, "the relations between attitudes and behaviors tend to increase in magnitude as the attitudinal and behavioral entities come to correspond more closely in terms of their target and action elements." (p. 911) The insight from these studies is that unless the attitudes being measured are relevant to the behavioral changes targeted, the measurements will not provide an adequate indicator of the training's value.

What if the targeted behavioral change is only vaguely defined? What if the attitudinal response cannot be solicited directly without confusion? The behavioral changes identified by Richard Farson (1996) as "new ways of thinking, new perspectives" do not evoke obvious behavioral targets or attitudinal measures. The actual benefits from management training and development programs are unapparent. There are no obvious observable target behaviors that result from an individual seeing things differently and thinking differently. Thus, an instrument that captures abstractly the change that occurs in the individual attitude is the best chance of identifying the potential behavior change in the individual. This paper presents a research measurement methodology capable of capturing this abstract attitude-behavior relationship.

The empirical research methodology is tested on the participants in a management-training program that incorporates a business simulation exercise called "Cycloan," developed by PriSim Business War Games Incorporated (Lefebvre 1999). This simulation exercise claims to produce the specific outcomes identified as seeing things differently and thinking differently.

GENERAL BACKGROUND AND THEORY

"All correct reasoning is a grand system of tautologies, but only God can make direct use of that fact. The rest of us must painstakingly and fallibly tease out the consequences of our assumptions" (Simon, 1996/1998, p. 15).

Proponents of business simulation exercises claim, "One of the most powerful benefits of simulation is that it changes in a variety of ways the perspectives of the managers who participate" (Reibstein & Chussil, 1997, p. 409). Case studies do seem to support this claim that business war games result in participants doing things differently, thinking longer term, seeing the big picture and better understanding the complexities of the competitive landscape (Gwynne, 1995; Hequet, 1995; Lefebvre, 1997; McCune, 1998; McIlvaine, 1999; Reibstein & Chussil, 1997; Sherman, 1996; Stewart, 1997; Wilson & Condom, 1995). Unfortunately investigators who have a vested interest in promoting the technique may bias many of these cases. Formal studies that actually attempt to measure the benefits of a simulation exercise are limited and have produced mixed results (Anderson & Lawton, 1992; Keys & Wolfe, 1990; Wolfe, 1985; Randel, et.al 1992; Chapman & Sorge 1999). This experimental study is the first of a series to actually quantify the impact of a business war game exercise intervention on a decision maker's business perceptions.

THE SETTING

This initial experimental study is conducted using a relatively homogeneous group of twenty-one senior vice presidents from a medium-sized service business. These participants all have a high level of understanding of their particular industry's dynamics and are familiar with the decisions that drive business profitability. All participants perform the same job function for the company and a have similar responsibilities. These participants are divided into four groups by the executive management.

The intervention is the custom business war game "Cycloan," which provides a competitive rehearsal simulation environment for the teams of participants to run the branch office of a service company. The decisions made by participant teams are those that are typically made in the management of a branch office and include the key drivers of business success. Performance in any particular year is based on a model of typical industry dynamics and the competitive landscape created by other participant teams. Facilitators (non-participants) are responsible for assuring that the teams are engaging in the conversations that are appropriate for making the simulated branch's decisions. These facilitators provide guidance and focus to the team conversations.

THE VALUE

This study provides valuable insight into what happens to participants in a business war game exercise. Based on experience, several researchers have documented the benefits of these exercises (Chapman & Sorge, 1999; Lefebvre, 1997; Reibstein & Chussil, 1997). The most common among these reported benefits include: practice in an environment without risk, increased creativity, decreased decision time, more focused competitive analysis, and increased cross-functional understanding. Although case studies identified previously and practitioner experience support these benefits, little empirical evidence is offered in the training literature on the change that an individual participant experiences. Thus, the question, whether or not participants in a business simulation exercise change the way that they act in their actual business environment, remains unanswered. Do the benefits listed above result in a quantifiable change in the individual? The answer to this question is not self-evident.

A first step, in addressing this difficult question is measuring whether or not a change has taken place in the individual. Has the business war game in some way changed the perspective of participants? As Evans and Wurster (Evans & Wurster, 2000) point out, the biggest risk today for businesses is not "legacy assets" but the "legacy mindset" (p. 66). In a landmark book, "Creativity"

(Csikszentmihalyi, 1996), Mihaly Csikszentmihalyi reported the results of interviewing over 90 of arguably the most creative people in the world. One of Csikszentmihalyi's major findings was that creative people "look at problems from as many viewpoints as possible" (p. 365). Great discoveries like Einstein's theory are the result of thinking differently (Sherman & Schultz, 1998, p. 235). These thinkers not only change their own perspectives, but they create a new perspective for others to follow. If the perspectives of people can be changed, all indications are that individuals will think differently, and creative new approaches will emerge. Thus, changing individual perspectives should be an objective of management training, and measuring this change should be the primary indicator of value.

A goal of this study is to develop the methodology necessary to measure the value of management training using an intervention specifically designed to change people's perspectives. This is done by implementing the semantic differential technique developed by Osgood, Suci, and Tannenbaum (1957). First, the study identifies changes in individual characterization of decision problems along several key dimensions, which are identified by semantic descriptors. These semantic descriptors are than measured to verify that the intervention has in fact resulted in a significant change. Using another set of semantic descriptors, the study identifies whether the individual's perception of which approach to take toward a specific decision has changed as the result of the intervention. The result is a study providing empirical evidence that a business war game exercise changes participants' perceptions of decisions and the way that they anticipate responding to those decisions.

A firm wishing to justify the capital expenditure of such a management development activity can point to empirical evidence that a measurable change in the participants has occurred. Using the proposed instrument and methodology, a firm will also be able to pinpoint exactly which dimensions have experienced the most significant change and verify that desired objectives have been achieved. If change has not been achieved along the desired dimensions, modification can be made to perhaps focus the development activity at specific problems and dimensions. Thus, the instrument provides a tool for identifying, measuring, communicating and targeting change in an organization's people.

THE OBJECTIVES

The experimental objective is to provide evidence for accepting a number of a-priori hypotheses focusing on the concept that participants completing a business war game exercise will characterize their decisions differently after the exercise and will identify different approaches to these same decisions. Generically, these hypotheses are presented as follows (where, Xn represents the decision targets, C corresponds to the decision characteristic measure, and A denotes the decision approach measure):

H1: Decision target [Xn] is characterized as ordinally more/less [C] after participation in the business simulation exercise.

H2: Decision target set [X1, X2, X3, X4, X5, X6] is reordered along the [C] dimension after participation in the business simulation exercise.

H3: Decision problem [Xn] is approached ordinally more/less [A] after participation in the business simulation exercise.

H4: Decision target set [X1, X2, X3, X4, X5, X6] is reordered along the [A] dimension after participation in the business simulation exercise.

Since the target decisions have not yet been defined in this paper, the specifics of these hypotheses will be developed and detailed in the next several sections. The experimental objective is simply stated as finding the supporting evidence to accept these hypotheses.

RESEARCH DESIGN

As the title of this paper implies, this study is designed for an experimental implementation. The complex intervention, called a business war game exercise, is designed to influence the two independent variables in this study, the decision characteristic and decision approach concepts. While the independent variable is changed by the intervention, the measurement instrument attempts to confirm the hypothesized changes in the dependent variables. In this design the dependent variables are eight measurements (semantic differential scales) associated with each decision concept.

The term experimental design has been used loosely up until this point. There are actually many different experimental designs described in the literature (Campbell & Stanley, 1963). These designs vary widely in their ability to control for the parameters that might influence the relationship between the independent and dependent variables. This study uses a pre-experimental one-group pretest-posttest design.

Only a single homogeneous group is subjected to the pretest measurement, business war-game intervention and posttest measurement sequence suggested by the design. Limiting the design to the pre-experimental grouping becomes a necessity when the target decisions are customized to the group being tested and the population is too small to allow for a reasonable control group. Thus, the a-priori hypotheses, H1 through H4, results may not be generalized easily to other groups. The intent however, is that this methodology will be replicated with other groups as part of an ongoing research effort.

Care is taken to ensure that the design is applied to maximize its power to measure the hypothesized relationships. Given the pre-experimental design, efforts are focused on establishing high levels of internal validity. The group is relatively homogeneous from an experience and

education level, allowing for an instrument design like the semantic differential, which relies on the participants having a common language set. Group homogeneity along with the small population receiving the intervention, only twenty-one (21) participants, provides some assurance that all participants experience the intervention equivalently.

The format chosen for the business war game exercise requires the participants to be away three days from their job demands, and focus their energy on running a simulated business. This format is ideal for controlling for many of the parameters that the passage of time can present. The pretest measurement is taken at the beginning of the first day, while the posttest measurement is taken at the end of the third day. In the interim time, the participants are staying in a remote location and instructions are given to the participants by the executive management that "running the simulated business should be considered their top priority." This time compressed, focused format weakens the impact that external factors may have on the participants and increases the experiment's validity.

METHODOLOGY

The target decision concepts are selected considering both the common decision set for the participants and the ability to rate these decisions along a variety of semantic scales. As an individual obtains more and more experience with a particular decision, a concept emerges. This process is exemplified by the learning of Hull's Chinese characters (Hull, 1920). By selecting common decisions, or decisions with which the participants are familiar, participants are able to recall an unambiguous mental concept that they can then use for rating the semantic differential scales. This reduces the major concern that participants will be rating the semantic scales based on different concepts, and increases the likelihood that the data can reliably be aggregated to generate the true meaning of the concept.

Another consideration is that the decisions selected are those also made during the business war game exercise. Since the intervention being tested, a custom business war game exercise, is designed with a limited number of decisions that focus participant learning, the set of possible decision concepts is constrained. The decisions included in a custom exercise are typically those that are identified by the designer consultants as "critical" to the running of the business. These critical decisions are the ones targeted for change by the business war game designers and therefore are the ones selected for measurement (Goosen 2001).

Figure 1 contains a short-list of concepts selected as potential candidates for measurement. Interviews with the business war game exercise designers and sponsors reduce this short-list to the six concepts highlighted in boxes. These six concepts cover a large range of business decisions, from personnel management to business strategy, and span the common decision set of the participants.

FIGURE 1: TARGET DECISION CONCEPTS

BUILDING Competencies	FORECASTING Sales
SETTING planned sales volume	MANAGING Employee Turnover
ESTABLISHING Strategy	TARGETING Customers
HIRING a Loan Officer	SELECTING Measures of Business Success
ALLOCATING Loan Officer Time	SELECTING Tactical Focus
ALLOCATING Area Sales Manager (ASM) Time	RESPONDING to Competitor Actions
TRAINING Loan Officers	SOURCING Leads

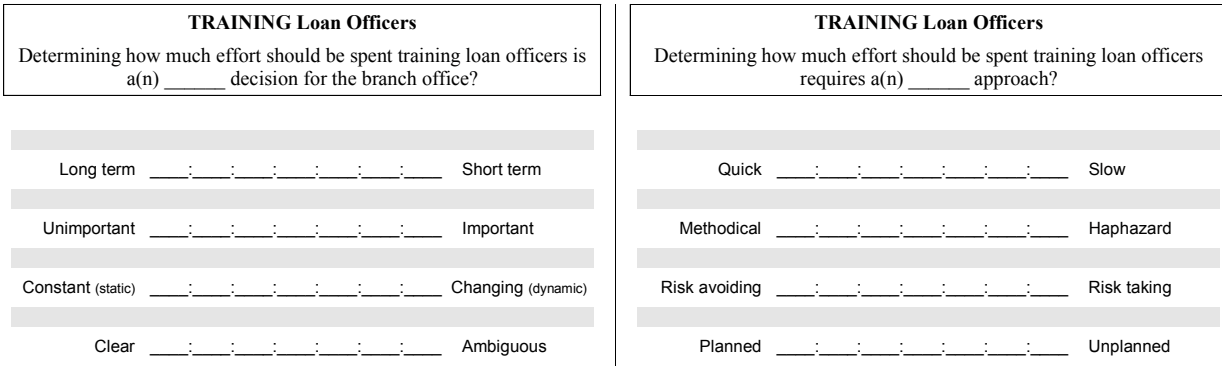
The six target, boxed, decision concepts selected from Figure 1 are abbreviated as follows: Strategy (establishing strategy), Hiring (hiring a loan officer), Time Block or Time Blocking (allocating loan officer time), Training (training loan officers), Targeting (targeting customers) and Sourcing (sourcing leads). These contractions are used in the remainder of this paper.

QUESTIONNAIRE DESIGN

A number of exploratory and pilot studies were conducted to develop this final instrument design. Two sets (characterization and approach) of eight bipolar semantic differential scales are developed as the proxy measurements for the each of the decision targets in Figure 1. Each scale is constructed in the standard seven-point rating format as described by Osgood et al. (1957). The result is a questionnaire instrument that collects data on the perceived magnitude and direction of the change in each decision problems characterization and selected approach.

The decision dimension questionnaire begins with a page of detailed instructions, describing the correct marking of the instrument. As suggested by a pilot study, the questions are grouped first by decision concept, with the decision characteristic concept in the left column and the decision approach concept in the right column. Across columns, the target decisions are presented so that the same decision-target that appears in the decision characteristic column is mirrored in the decision approach column. Finally, space limitations allow only two decision targets to be included on each page. This layout is illustrated in Figure 2. Decision targets and measurement scales are ordered to reduce respondent bias (Emory 1985).

FIGURE 2: EXPERIMENTAL FIELD STUDY – QUESTION/LAYOUT



A-PRIORI HYPOTHESIS FORMULATION

An a-priori hypothesis can be formulated for each target decision and semantic differential scale combination, for a total of ninety-six hypotheses of the form H1 and H3. The construction of each hypothesis is based on the objectives of the business war game exercise. For example, the assumption is made a-priori that the target decision "establishing strategy" will be characterized as more "ambiguous" and will be identified as requiring a more "gut" approach. The objective of the business war game exercise is to identify applicable strategic techniques that will change the perspective on the target decision toward "clear" and the participant's thinking toward "textbook." Several of the highly targeted changes are identified in Figure 3.

FIGURE 3: A-PRIORI HYPOTHESES FOR "TARGETED" CHANGES

Independent Variable (Decision Target) X	Dependent Variable C	Direction of movement* < , > <+>, <0>	Raison d'être
Characteristic			
Hiring	Small-Big	>	hiring significantly impacts BWG performance
Training	Small-Big	>	training significantly impacts BWG performance
Time Blocking	Short Term-Long Term	>	competencies deteriorate with time and build with time spent
Strategy	Clear-Ambiguous	<	applicable strategy techniques identified in BWG
Targeting	Constant-Changing	<	BWG techniques increase understanding of dynamics
Sourcing	Constant-Changing	<	BWG techniques increase understanding of dynamics
Approach			
Hiring	Detailed-Big Picture	>	hiring decision made in context rich environment
Training	Detailed-Big Picture	>	training decision made in context rich environment
Time Blocking	Detailed-Big Picture	>	time blocking decision made in context rich environment
Strategy	Textbook-Gut	<	key strategic planning tools identified in BWG
Targeting	Detailed-Big Picture	>	targeting decision made in context rich environment
Sourcing	Detailed-Big Picture	>	sourcing decision made in context rich environment

Notes	* Key	
* "Direction of Movement" - is the expected direction of change along the semantic differential scale (C) as indicated. For example, a direction ">" along the "Small-Big" scale would imply that the decision is expected to be perceived by the respondent as "Bigger" after the BWG	<	significant movement toward left pole
	>	significant movement toward the right pole
	<+>	significant movement but no direction specified
	<0>	no change anticipated
	BWG	Business War Game exercise

The first set of hypotheses, H1 and H3, attempt to confirm that a change in decision perception and a change in thinking has occurred in the war game exercise participants. This confirmation is one indicator that the participant's reaction in an actual decision situation will be different as the result of the exercise. As recognized previously, prior research indicates that a connection exists between perception and response (Csikszentmihalyi, 1996; Evans & Wurster, 2000; Sherman & Schultz, 1998). In the context of this study, the interest is in the relationship between perspective and decision, and using a business war game exercise to improve the decision making of the participants.

A study conducted by Klein (1993) analyzes data from different domains and more than six hundred decision points

to identify the sources of decision errors. Three sources emerge from these data; lack of experience, lack of information, and explaining away. The third source, explaining away, is the result of a perception paradigm. The decision-maker has a mental map of the decision that is difficult to change. This study attempts to measure the change in the decision-maker's mental map using the semantic differential technique. Verification of this change is the focus of the first set of hypotheses, H1 and H3.

Only sixteen hypotheses can be generated from H2 and H4. In this case, the pre-test data will be used as the basis for the a-priori ordering and it will be compared to the post-test data to validate the hypotheses acceptance. The second set of hypotheses, H2 and H4, focus on the participants' decision set. If the ranking within the decision set has changed, intuition indicates that the priority the decision-maker places on decisions within that decision set also changes. This intuition is supported in the literature by studies in naturalistic decision-making (Klein, 1998; Zsambok, 1997). Thus, being able to confirm that a change has occurred in the decision-maker's decision perception can indicate that future decisions will be made using different priorities.

Field and intervention constraints limit the study to decisions that are considered "important" to the executive management of the participating firm. This does not imply that the participants will actually rate the decisions as "important;" however, it is a logical assumption. Thus, the study's interpretation relies on only intuitive support that decisions rated as "unimportant" will fit the same model. This interpretation is aided by the semantic differential technique that measures different magnitudes of importance. Even though all the decisions may be rated as "important," there exist several levels of "important" that are accounted for in the data analysis and model verification.

DATA PREPARATION

After verifying completeness, responses are coded and labeled as indicated in Figure 4. For example, a question is labeled "H S/C -," to quickly communicate that the decision target is "Hiring," the semantic scale is the decision characteristic scale "simple-complex" and the scale's origin is the pretest questionnaire. As suggested by Osgood et al. (1957), the coding is from "-3" to "+3", with the central point on the scale designated as "0" to represent neutrality. Using this coding, it becomes relatively easy to identify both the polar-direction, by the number's sign, and the polar-magnitude, by the absolute value of the score.

FIGURE 4: QUESTIONNAIRE CODING AND LABELING KEY

Decision Targets	Label	CODE
HIRING a Loan Officer	*Hiring*	**H**
TRAINING Loan Officers	*Training*	**TR**
ALLOCATING Loan Officer Time	*Time Blocking*	**TB**
ESTABLISHING Strategy	*Strategy*	**ST**
SOURCING Leads	*Sourcing*	**SO**
TARGETING Customers	*Targeting*	**TA**

Questionnaires	CODE
Pretest	**-**
Posttest	**+**

Ordinal/Interval Label	Score (Left or Right)
extremely	**-3 or +3**
quite	**-2 or +2**
slightly	**-1 or +1**
equally	**0**

Characteristic Scales	CODE
Simple - Complex	**S/C**
Short Term-Long Term	**S/L**
Reversible - Irreversible	**R/I**
Unimportant-Important	**U/I**
Low Risk-High risk	**L/H**
Constant-Changing	**C/C**
Small-Big	**S/B**
Clear-Ambiguous	**C/A**

Approach Scales	CODE
Textbook-Gut	**T/G**
Quick-Slow	**Q/S**
Active-Passive	**A/P**
Methodical-Haphazard	**M/H**
Individual-Team	**I/T**
Risk Avoiding-Risk Taking	**R/R**
Detailed-Big Picture	**D/B**
Planned Unplanned	**P/U**

In this study, there are no cases of missing data. Thus, data editing focuses on identifying any respondents that appeared to mark the scales in an arbitrary manner. Constraints did not allow for the inclusion of test scales in the experimental design. Therefore, editing relies on the subjective inspection of the questionnaires to identify cases where a particular scale is marked at the same level for every decision target, or a long sequence of scales is marked at the same level. No abnormalities are noted in the data set. A total of twenty-one matching, pretest and posttest, questionnaires are collected, representing a one hundred percent response rate.

DATA ANALYSIS AND FINDINGS FOR CONCEPT CHANGE (H1 AND H3)

In this analysis, support is sought for the a-priori hypotheses H1 and H3. Basic data analysis techniques are used to assemble this evidence. Additional support for these hypotheses, and evidence supporting H2 and H4 requires a

more sophisticated analysis. A number of statistical tools are used to analyze the data. The analysis begins with some simple summary statistics based on the raw data tables. This is followed by an analysis of the gap data, which is the difference between the pretest and posttest measurements. Finally, the matched pairs t-test and Wilcoxon signed-rank test are performed to assess significance. For brevity, the summary statistics and data presented in this section will be restricted to the twelve measurements identified in Figure 3, which are called the "targeted twelve."

Examining the data prior to subjecting it to more sophisticated techniques provides the researcher with critical insights into the characteristics of the data set. Each variable, or semantic differential scale is examined using a frequency histogram, standard descriptive statistics (mean, median, standard deviation, minimum, etc.) and plots testing for normality (normal probability plots and rootograms). Table 1 and Table 2 provide a results summary, limited to the "targeted twelve," for the pretest and posttest data respectively. These tables are read by first selecting a scale, using the key in Figure 4 to identify what scale is being tested and reading the summary data across each row. For example, the first scale in Table 1 is H S/B -, which is decoded as the pretest small/big scale for the hiring decision characterization. The number of data points is 21 with a minimum value of –2, a maximum value of 3, a mean of 1.90 indicating that participants view the hiring decision as quite big. Normality is checked to verify that the more advanced statistical test used in later analysis are valid.

TABLE 1: TARGETED TWELVE PRETEST SUMMARY STATISTICS

Scale	Count	Minimum	Maximum	Mean	Ordinal Descriptor	Median	Mode	Standard Deviation	Skewness	Kurtosis	Normal Probability	Rootogram
"Decision Characteristic" Concept												
H S/B -	21	-2	3	1.90	quite+	2	2	1.18	-2.02	5.42	Fail	Pass
TR S/B -	21	-2	3	1.71	quite+	2	2	1.35	-1.46	2.04	Fail	Pass
TB S/L -	21	-2	3	0.76	slightly+	1	2	1.79	-0.24	-1.39	Fail	Fail
ST C/A -	21	-3	2	-0.86	slightly-	-1	-3	1.93	0.33	-1.44	Fail	Fail
SO C/C -	21	1	3	2.33	extreme+	2	2	0.66	-0.47	-0.55	Fail	Pass
TA C/C -	21	-2	3	1.90	quite+	2	3	1.26	-1.62	3.43	Fail	Pass
"Decision Approach" Concept												
H D/B -	21	-3	3	0.05	slightly+	0	2	1.80	-0.08	-1.40	Fail	Pass
TR D/B -	21	-3	3	-0.48	slightly-	0	-2	1.83	0.42	-1.02	Fail	Fail
TB D/B -	21	-3	2	-0.81	slightly-	-1	-2	1.63	0.65	-0.85	Fail	Fail
ST T/G -	21	-3	3	-0.67	slightly-	-1	-1	1.35	0.79	1.37	Fail	Pass
SO D/B -	21	-3	3	0.43	slightly+	1	2	1.94	-0.31	-1.47	Fail	Fail
TA D/B -	21	-3	3	0.38	slightly+	1	2	2.13	-0.25	-1.48	Fail	Fail

TABLE 2: TARGETED TWELVE POSTTEST SUMMARY STATISTICS

Scale	Count	Minimum	Maximum	Mean	Ordinal Descriptor	Median	Mode	Standard Deviation	Skewness	Kurtosis	Normal Probability	Rootogram
"Decision Characteristic" Concept												
H S/B +	21	1	3	2.14	extreme+	2	2	0.65	-0.14	-0.43	Fail	Pass
TR S/B +	21	1	3	2.19	extreme+	2	3	0.87	-0.40	-1.61	Fail	Fail
TB S/L +	21	-1	3	1.57	quite+	2	2	1.12	-0.67	-0.12	Fail	Pass
ST C/A +	21	-3	1	-1.67	quite-	-2	-3	1.28	0.57	-0.87	Fail	Fail
SO C/C +	21	-3	3	1.33	quite+	2	2	1.77	-1.34	0.94	Fail	Pass
TA C/C +	21	-3	3	1.19	quite+	2	2	1.94	-1.02	-0.21	Fail	Fail
"Decision Approach" Concept												
H D/B +	21	-3	3	1.00	slightly+	2	2	1.90	-0.92	-0.44	Fail	Pass
TR D/B +	21	-3	3	0.62	slightly+	2	2	2.16	-0.44	-1.60	Fail	Fail
TB D/B +	21	-3	3	0.52	slightly+	2	2	2.16	-0.40	-1.46	Fail	Fail
ST T/G +	21	-3	2	-1.10	quite-	-1	-2	1.22	0.56	0.49	Fail	Pass
SO D/B +	21	-3	3	0.90	slightly+	1	3	2.05	-0.52	-1.15	Fail	Fail
TA D/B +	21	-2	3	1.19	quite+	2	2	1.54	-0.63	-0.78	Fail	Pass

The raw data are then transformed into a gap data set by subtracting the pretest measures from the posttest measures. These data are subjected to the same descriptive statistical analysis as the original data. The gap summary data for the targeted twelve are presented in Table 3.

TABLE 3: TARGETED TWELVE GAP SUMMARY STATISTICS

Scale	Count	Minimum	Maximum	Mean	Ordinal Descriptor (Mean)	Median	Mode	Standard Deviation	Skewness	Kurtosis	Normal Probability Plot	Rootogram
"Decision Characteristic" Concept												
H S/B	21	-1	4	0.24	+	0	0	1.26	1.82	3.73	Fail	Pass
TR S/B	21	-1	3	0.48	+	0	0	1.08	0.73	0.08	Fail	Pass
TB S/L	21	-4	4	0.81	+	1	0	1.86	-0.21	1.43	Fail	Fail
ST C/A	21	-5	3	-0.81	-	0	0	2.11	-0.49	-0.12	Pass	Pass
SO C/C	21	-6	1	-1.00	-	0	0	1.87	-1.52	1.77	Fail	Fail
TA C/C	21	-6	4	-0.71	-	0	0	2.12	-0.66	1.65	Fail	Fail
"Decision Approach" Concept												
H D/B	21	-3	6	0.95	+	0	0	2.52	0.25	-0.36	Pass	Pass
TR D/B	21	-2	6	1.10	+	0	0	2.12	1.05	0.16	Fail	Fail
TB D/B	21	-5	6	1.33	+	1	4	2.83	-0.49	-0.13	Pass	Pass
ST T/G	21	-3	2	-0.43	-	0	0	1.47	-0.10	-0.71	Pass	Pass
SO D/B	21	-4	6	0.48	+	0	0	2.40	0.40	0.78	Pass	Pass
TA D/B	21	-4	6	0.81	+	0	0	2.40	0.22	0.30	Fail	Fail

The first indication of change in meaning for the decision concepts comes by observing the gap data. If a gap is greater than zero in absolute value then a change has occurred. The question is at what value greater than zero is the change meaningful. In other words, at what gap value is the statement confirmed that the concept's meaning has changed? These questions can be answered using the parametric matched t-test on the raw pretest and posttest data files. However, since much of the data fails the normality tests, the non-parametric Wilcoxon signed-rank test will also be conducted for additional support using the gap data.

The matched t-test of the means, tests the hypothesis by defining the null hypothesis H0: mean difference = pretest mean – posttest mean = 0 (two tailed), versus the alternative hypothesis H1: mean difference [not =] pretest mean – posttest mean [not =] 0 (two tailed). A "one-tailed" matched pair t-test is also performed based on the directional a-priori hypotheses identified previously. The "one-tailed" test defines the null hypothesis H0: mean difference >= 0 (or <=0), verses the alternative hypothesis H1: mean difference < 0 (or > 0). These results are tabulated for the targeted twelve in Table 4. The table is keyed for the following confidence levels: 95% are double underlined and in bold, 90% are single underlined and in bold, and 80% are single underlined and in italics.

TABLE 4: HYPOTHESIS RESULTS (MATCHED PAIRED T-TEST)

p-value Key

(=)	Reject H0: Posttest-Pretest = 0; Accept H1: Posttest -Pretest not = 0
(>)	Reject H0: Posttest-Pretest >= 0; Accept H1: Posttest-Pretest < 0
(<)	Reject H0: Posttest-Pretest <= 0; Accept H1: Posttest-Pretest > 0
*	a priori hypothesis

Scale	N	Degrees of freedom	Estimated Mean Difference	p-value (=)	p-value (<)	p-value (>)
"Decision Characteristic" Concept						
H S/B	21	20	0.2	0.40	0.80	*0.20* *
TR S/B	21	20	0.5	**0.06**	0.97	**0.03** *
TB S/L	21	20	0.8	**0.06**	0.97	**0.03** *
ST C/A	21	20	-0.8	**0.09**	**0.05** *	0.95
SO C/C	21	20	-1.0	**0.02**	**0.01** *	0.99
TA C/C	21	20	-0.7	*0.14*	**0.07** *	0.93
"Decision Approach" Concept						
H D/B	21	20	1.0	**0.10**	0.95	**0.05** *
TR D/B	21	20	1.1	**0.03**	0.99	**0.01** *
TB D/B	21	20	1.3	**0.04**	0.98	**0.02** *
ST T/G	21	20	-0.4	*0.20*	**0.10** *	0.90
SO D/B	21	20	0.5	0.37	0.81	*0.19* *
TA D/B	21	20	0.8	*0.14*	0.93	**0.07** *

Alternatively, the one-sample Wilcoxon signed-rank test of the median, tests the hypothesis by defining the null hypothesis H0: median gap = hypothesized median gap = 0, versus the alternative hypothesis H1: median gap [not =] hypothesized median gap [not =] 0. These results are tabulated for the targeted twelve in Table 5. The table is keyed for the following confidence levels 95% are double underlined and in bold, 90% are single underlined and in bold, and 80% are single underlined and in italics.

TABLE 5: HYPOTHESIS RESULTS (WILCOXON SIGNED-RANK TEST)

p-value Key	
(=)	Reject H0: Gap = 0; Accept H1: Gap not = 0
(>)	Reject H0: Gap >= 0; Accept H1: Gap < 0
(<)	Reject H0: Gap <= 0; Accept H1: Gap > 0
*	a priori hypothesis

Scale	N	N for Test	Wilcoxon Statistic	Estimated Median Gap	p-value (=)	p-value (<)	p-value (>)
"Decision Characteristic" Concept							
H S/B	21	10	32.5	0.0	0.65	0.71	0.32 *
TR S/B	21	11	54	-0.5	**0.07**	0.97	**0.03** *
TB S/L	21	14	84.5	-0.5	**0.05**	0.98	**0.02** *
ST C/A	21	14	26.5	0.5	*0.11*	**0.06** *	0.95
SO C/C	21	10	5	0.5	**0.03**	**0.01** *	0.99
TA C/C	21	10	13	0.5	*0.15*	**0.08** *	0.94
"Decision Approach" Concept							
H D/B	21	14	78.5	-1.0	*0.11*	0.95	**0.06** *
TR D/B	21	9	41	-1.0	**0.03**	0.99	**0.02** *
TB D/B	21	17	118	-1.5	**0.05**	0.98	**0.03** *
ST T/G	21	16	44	0.5	0.22	*0.11* *	0.90
SO D/B	21	16	84.5	-0.5	0.41	0.81	0.20 *
TA D/B	21	12	57.5	-0.5	*0.16*	0.93	**0.08** *

The gap data can also be used to identify specific changes in individuals. A non-statistical heuristic measure is simply the frequency, or count of values exceeding certain changes. Snider and Osgood (1969) states that "the evidence shows that for individual subjects a shift of more than two scale units probably represents a significant change or difference in meaning" (p. 79). They go on to add, "for group data ("cultural meanings"), changes or differences in measured meaning as small as one-half of a scale unit are significant at the 5 percent level" (p. 79). Given the homogeneity of the group and the specificity of the concepts, it can be argued that the data collected in this study is "group data" or at least somewhere between the two extremes outlined by Snider and Osgood (1969). Thus, this study will conservatively consider all changes that equal or exceeds two scale units as "significant" when analyzing individual cases. Table 6 contains these data for the targeted twelve semantic differential scales.

TABLE 6: INDIVIDUAL CASES – SUMMARY OF CHANGE

Scale	Hypothesized Direction	Count with significant change (>= 2 units)	Percentage with significant change (>= 2 units)	Count with significant change in direction hypothesized (>= 2 units)	Percentage with significant change in direction hypothesized (>= 2 units)
"Decision Characteristic" Concept					
H S/B	>	2	9.5%	2	9.5%
TR S/B	>	4	19.0%	4	19.0%
TB S/L	>	7	33.3%	6	28.6%
ST C/A	<	9	42.9%	7	33.3%
SO C/C	<	6	28.6%	6	28.6%
TA C/C	<	6	28.6%	5	23.8%
"Decision Approach" Concept					
H D/B	>	10	47.6%	7	33.3%
TR D/B	>	8	38.1%	7	33.3%
TB D/B	>	13	61.9%	10	47.6%
ST T/G	<	7	33.3%	5	23.8%
SO D/B	>	8	38.1%	5	23.8%
TA D/B	>	9	42.9%	6	28.6%

DATA ANALYSIS AND FINDINGS FOR ORDERING CHANGE (H2 AND H4)

Assessing whether or not a change has occurred in the perceptual rank ordering of the decision targets can be done with the same data set. This process provides the first supporting evidence for the hypotheses H2 and H4, which claim a-priori that a change has occurred. The method uses a derived measurement since data is not collected directly on the participant's perceptual ordering of the decision targets.

Statistical tools, which require the assumption of interval data, are again used in the assessment. The mean and t-test statistics are chosen to evaluate the magnitude of the change and whether the change is significant. Due to the large number of t-test evaluations necessary to determine which of the changes are significant, there is no easy method for indicating these findings graphically, thus, Table 7 is used in summary.

TABLE 7: DECISION CHARACTERISTIC CONCEPT – H2

Semantic Scale		1	2	3	4	5	6	Signficance
Simple / Complex	**Pretest**	Training	Time Block	Hiring	Targeting	Sourcing	Strategy	
	Posttest	Time Block	Training	Sourcing	Targeting	Hiring	Strategy	None at 80% Confidence
Long Term / Short Term	**Pretest**	Time Block	Targeting	Sourcing	Training	Strategy	Hiring	
	Posttest	Sourcing	Targeting	Time Block	Hiring	Training	Strategy	None at 80% Confidence
Reversible / Irreversible	**Pretest**	Targeting	Time Block	Sourcing	Strategy	Hiring	Training	
	Posttest	Sourcing	Targeting	Time Block	Strategy	Hiring	Training	None at 80% Confidence
Unimportant / Important	**Pretest**	Time Block	Strategy	Targeting	Sourcing	Training	Hiring	
	Posttest	Targeting	Time Block	Sourcing	Strategy	Training	Hiring	None at 80% Confidence
Low Risk / High Risk	**Pretest**	Strategy	Targeting	Hiring	Sourcing	Time Block	Training	
	Posttest	Hiring	Strategy	Targeting	Sourcing	Time Block	Training	None at 80% Confidence
Constant / Changing	**Pretest**	Training	Hiring	Time Block	Targeting	Strategy	Sourcing	
	Posttest	Targeting	Training	Strategy	Time Block	Sourcing	Hiring	Hiring and Targeting 87% confidence
Small / Big	**Pretest**	Time Block	Training	Hiring	Sourcing	Targeting	Strategy	
	Posttest	Time Block	Strategy	Targeting	Hiring	Training	Sourcing	None at 80% Confidence
Clear / Ambiguous	**Pretest**	Training	Sourcing	Targeting	Hiring	Time Block	Strategy	
	Posttest	Training	Strategy	Targeting	Hiring	Time Block	Sourcing	None at 80% Confidence

The same procedure is used to assess ordering changes in the decision approach concept. Tests are conducted to determine significance, and a summary is provided in Table 8.

TABLE 8: DECISION APPROACH CONCEPT – H4

Semantic Scale		1	2	3	4	5	6	Signficance
Textbook / Gut	**Pretest**	Training	Targeting	Strategy	Sourcing	Time Block	Hiring	
	Posttest	Training	Targeting	Strategy	Sourcing	Hiring	Time Block	None at 80% Confidence
Quick / Slow	**Pretest**	Time Block	Targeting	Training	Sourcing	Hiring	Strategy	
	Posttest	Targeting	Time Block	Training	Hiring	Sourcing	Strategy	None at 80% Confidence
Active / Passive	**Pretest**	Hiring	Sourcing	Time Block	Strategy	Training	Targeting	
	Posttest	Hiring	Training	Strategy	Sourcing	Targeting	Time Block	None at 80% Confidence
Methodical / Haphazard	**Pretest**	Training	Strategy	Targeting	Sourcing	Hiring	Time Block	
	Posttest	Strategy	Training	Sourcing	Time Block	Targeting	Hiring	None at 80% Confidence
Individual / Team	**Pretest**	Time Block	Sourcing	Hiring	Targeting	Training	Strategy	
	Posttest	Time Block	Hiring	Training	Targeting	Sourcing	Strategy	None at 80% Confidence
Risk Avoiding / Risk Taking	**Pretest**	Training	Time Block	Hiring	Strategy	Targeting	Sourcing	
	Posttest	Training	Time Block	Sourcing	Targeting	Strategy	Hiring	None at 80% Confidence
Detailed / Big Picture	**Pretest**	Time Block	Training	Hiring	Targeting	Sourcing	Strategy	
	Posttest	Time Block	Training	Sourcing	Hiring	Strategy	Targeting	None at 80% Confidence
Planned / Unplanned	**Pretest**	Hiring	Strategy	Training	Sourcing	Targeting	Time Block	
	Posttest	Sourcing	Strategy	Training	Hiring	Time Block	Targeting	None at 80% Confidence

DISCUSSION

Identifying that significant change has occurred on these specific semantic differential scales, as the result of participating in a business war game, is a first step in understanding what has "really" changed. These results indicate that significant changes have occurred along targeted dimensions, confirming between 75% and 83% of the tests. Results from the un-targeted dimensions are less compelling, confirming roughly a third of the tests.

A total of 48 a-priori hypotheses of the form H1 were stated in the research, with 12.5% confirmed with at least 95% confidence, 22.9% confirmed with at least 90% confidence, and 35.4% confirmed with at least 80% confidence using the matched pair t-test. Equivalently, 48 a-priori hypotheses of the form H2 were tested using the matched pair t-test, confirming 12.5% with at least 95% confidence, 20.8% with at least 90% confidence, and 39.6% with at least 80% confidence.

The focus of this paper is on the targeted twelve hypotheses identified in Figure 3. The matched pair t-test confirms ten of the targeted twelve hypotheses, at the 90% confidence level, indicating that the study participant's perceptual characterization and selected approaches changed in an a-priori prescribed way. Specifically, the paired t-test confirmed five of the six desired decision characterization changes, with participant's characterizing the training decision as bigger, the time blocking decision as longer term, the strategy decision as clearer, the sourcing decision as less changing, and the targeting decision as less changing. Also confirmed were five of the six approach changes, with the test indicating that participants would take a bigger picture approach to the hiring, training, time blocking, and targeting decisions; and a less gut approach to strategy. Similar results are identified by the Wilcoxon signed-rank test, which confirms nine of the same decisions, failing to confirm only the hypothesis that participants would take a less gut approach to strategy.

Statistically significant support is not found for the two rank ordering hypotheses, H2 and H4. Although there is some evidence to indicate that a re-ordering of decision targets has occurred as indicated by visually comparing the results presented in Table 7 and 8. The only statistically significant change occurs along the constant-changing dimension with the hiring decision becoming more "dynamic" relative to the targeting decision, after the business war game exercise. Unfortunately, the data is not sufficient to make similar claims for the other decision targets.

RELIABILITY AND VALIDITY OF RESULTS

Reliability is typically associated with the concept of consistency. It is concerned with estimating the degree to which the measurement being taken is free from random or unstable error (Emory, 1985, p. 98). To assess the reliability of this experimental field study, two criteria are utilized, stability and equivalence. Stability is an indictor of how consistent the results would be if the measurements were repeated on the same person using the same instrument. This re-testing control is not implemented in the methodology because the time of posttest is critical to obtaining an accurate measurement. It is extremely difficult to control for the large number of confounding parameters,

which can contaminate the measurement with the passage of time. Therefore, stability within the sample group is an assumption based on the questionnaire's careful construction. The decision targets that are chosen for measurement all have high levels of familiarity among the participants. Thus, there is little reason to believe that the respondents will alter their responses with the passage of time and without an environmental change.

Another reliability concern is with the equivalence of test groups. In this study, only a single homogeneous test group is subjected to the measurement. As this study is repeated, the models developed for the decision characteristic and decision approach concepts need to be tested for equivalence.

Reliability is a necessary but not sufficient condition for achieving validity. Validity is the extent that the differences being measured by the instrument reflect the true differences among those being tested. Three major forms have been identified in the literature: content, criterion-related, and construct (Emory, 1985, p. 95). The first, content validity, is addressed by selecting six of the major decisions contemplated by the participants in the business war game exercise. This selection was determined by the exercise developers and the participants' executive management to cover a broad range of common business decisions.

The extensive scale development effort, highlighted in the decision dimension and pilot studies (Scherpereel, 2001) provide some assurance that the criteria are valid. The measures used for profiling the different decision targets seem to mirror the general description of these same decisions. Using multiple criterion to measure the same dimension in the semantic space provides further assurance that the measurement instrument has a high level of criterion-related validity.

The two features, multiple measurement scales and multiple decision targets, help increase the study's construct validity. The study is designed to develop and assess these constructs for their predictive abilities. Indication is that the models, or constructs, developed are consistent from decision target to decision target. For example, the decision targets "sourcing" and "targeting," which a-priori might be hypothesized as having similar profiles, are confirmed to have similar profiles by this study's data. Thus, it is shown that measurements on these devised scales correlate in predictable ways for different propositions, and construct validity is enhanced.

CONCLUSIONS AND FUTURE RESEARCH

Do people participating in a business war game exercise see things differently and think differently? This is the fundamental question answered "yes" by this research.

The techniques developed in this research are applied to a real world problem of evaluating the benefits of a "business war game" exercise. The decision measurement technique was able to clearly demonstrate that a "business war game" exercise changes the way decision-makers see decision problems and the way they think about these problems. The exercise effectiveness was measured along specific dimensions to verify that a decision-maker's perception changed according to the sponsor's objectives. Measuring the extent to which a business war game exercise is able to meet specific objectives, provides the first verifiable test of value available for this intervention. The technique's success provides quantitatively measured justification to a business leader sponsoring a business war game training activity in their organization.

The measurement methodologies developed specifically to quantify the value of a business war game exercise can be applied to other training and development activities. The methodologies developed in this research are especially applicable to executive development training, where training effectiveness is not easily quantified using traditional evaluation techniques. In cases where other techniques can be used, this research offers an alternative measurement instrument. The principal advantage of this new methodology is that it allows an immediate indication of the impact of an executive training and development activity. This contribution is significant for both decision-makers, who are attempting to justify expenditures on these programs, and firms, which are promoting the use of these programs.

The empirical evidence presented in this research is relatively limited. The sample size is small and is selected from a single homogeneous group of decision-makers. This limitation leaves open a number of "what if" questions. What if more decision-makers are studied, will the results be the same? What if different decision-makers are studied? What if the group studied is not homogeneous? By expanding the investigation to include a larger number of decision-makers with a more diverse makeup, the validity of the conclusions might be strengthened. This expansion will also allow more sophisticated techniques to be used to statistically model the data structure.

CITATIONS

Ajzen I., a. M. F. (1977). Attitude-behavior Relations: A Theoretical Analysis and Review of Empirical Research. *Psychological Bulletin*, 84, 888-918.

Alliger, G. M., Tannenbaum, S. I., & Jr., W. B. (1997). A Meta-analysis of the Relations among Training Criteria. *Personnel Psychology*, 50, 341-358.

Anderson, P. H., & Lawton, L. (1992). A Survey of Methods used for Evaluating Student Performance on Business Simulations. *Simulation and Games*, 23(4), 490-498.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago, Illinois: Rand McNally.

Chapman, K. J., & Sorge, C. L. (1999). Can a Simulation Help Achieve Course Objectives? An Exploratory Study Investigating Differences Among Instructional

Tools. *Journal of Education for Business*, 74(4), 225-230.

Csikszentmihalyi, M. (1996). *Creativity: Flow and the Psychology of Discovery and Invention*. New York: Harper Collins.

Day, G. S., & Reibstein, D. J. (1997). *Wharton on Dynamic Competitive Strategy*. New York: John Wiley and Sons, Inc.

Emory, C. W. (1985). *Business Research Methods*. Homewood, Illinois: IRWIN.

Evans, P., & Wurster, T. S. (2000). *Blown to Bits: How the New Economics of Information Transforms Strategy*. Boston, Massachusetts: Harvard Business School Press.

Fagiano, D. (1995). Making Training Quantifiable. *Supervision*, 56, 12-13.

Farson, R. (1996). *Management of the Absurd*. New York: Touchstone.

Filipczak, B., Picard, M., & Stamps, D. (1998). Measuring Training's Contribution to Intellectual Capital. *Training and Development*, 35, 14.

Goosen, K. R. (2001). Purpose and learning benefits of simulations: A design and development perspective. *Simulation and Gaming*, 32(1), 17.

Gwynne, P. (1995). Let the Battle Begin. *Across the Board*, 32, 33-37.

Hequet, M. (1995). Games that Teach. *Training*, 32, 53-58.

Hull, C. L. (1920). Quantitative Aspects of the Evolution of Concepts. *Psychology Monographs*, 38(123), 1-86.

Iaffaldano, M. T., & Muchinsky, P. M. (1985). Job Satisfaction and Performance. *Psychological Bulletin*, 97, 251-273.

Keys, B., & Wolfe, J. (1990). The Role of Management Games and Simulations in Education and Research. *Journal of Management*, 16, 307-336.

Kirkpatrick, D. J. (1994). *Evaluating Training Programs: The Four Levels*. New York: Berrett-Koehler.

Klein, G. (1993). A Recognition-Primed Decision (RPD) Model of Rapid Decision Making. In G. Klein & J. Orasanu & R. Calderwood & C. E. Zsambok (Eds.), *Decision Making in Action: Models and Methods*. Norwood, NJ: Ablex.

Klein, G. (1998). *Sources of Power: How People Make Decisions*. Cambridge, Massachusetts: The MIT Press.

Lefebvre, J. R. (1997). The Battle for Change. *Mortgage Banking*, January, 1-7.

Lefebvre, J. R. (1999). *Cycloan: a mortgage banking simulation.* Illinois: PriSim Business War Games Incorporated.

McCune, J. C. (1998). The Game of Business: managers are learning how to play in the real marketplace through computer simulations. *Management Review*, 87, 56-58.

McIlvaine, A. R. (1999). Games Employees Play. Human Resource Executive, February, 37-39.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The Measurement of Meaning*. Urbana, Illinois: The University of Illinois Press.

Randel, J. M., Morris, B. A., Wetzel, C. D., & Whitehill, B. V. (1992). The Effectiveness of Games for Educational Purposes: A Review of Recent Research. *Simulation and Games*, 23(3), 221-276.

Reibstein, D. J., & Chussil, M. J. (1997). Putting the Lesson Before the Test: Using Simulation to Analyze and Develop Competitive Strategies. In G. S. Day & D. J. Reibstein (Eds.), *Wharton on Dynamic Competitive Strategy* (pp. 395-423). New York: John Wiley and Sons, Inc.

Reingold, J. (1999). Exec Ed: Learning to Lead. *Business Week*, 42, 1-6.

Scherpereel, C. M. (2001). *Seeking the Dimensions of Decision-Making: An exploratory study*. Flagstaff: Northern Arizona University.

Shelton, S., & Alliger, G. M. (1993). Who's Afraid of Level 4 Evaluation? A Practical Approach. *Training and Development*, 47, 43-46.

Sherman, E. (1996). The Virtual Business. *Traffic World*, 247, 45-46.

Sherman, H., & Schultz, R. (1998). *Open Boundaries: creating business innovation through complexity*. Reading, Massachusetts: Perseus Books.

Simon, H. A. (1996/1998). *The Sciences of the Artificial*. Cambridge, Massachusetts: The MIT Press.

Snider, J. G., & Osgood, C. E. (Eds.). (1969). *Semantic Differential Technique: A Sourcebook*. Chicago, Illinois: Aldine Publishing Company.

Stewart, T. A. (1997). The Dance Steps get Trickier all the Time. *Fortune*, 135, 157-158.

Weigel, R. H., Vernon, D. T. A., & Tognacci, L. N. (1974). Specificity of the Attitude as a Determinant of Attitude-behavior Congruence. *Journal of Personality and Social Psychology*, 30, 724-728.

Wilson, J. R., & Condom, P. (1995). Entertainment Drives Simulation Technology. Interavia, *Business & Technology*, 50, 13-14.

Wolfe, J. (1985). The Teaching Effectiveness of Games in Collegiate Business Courses. *Simulation and Games*, 12(3), 23-40.

Zsambok, C. E. (1997). Naturalistic Decision Making: Where Are We Now? In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making: Where Are We Now?* (pp. 3-16). Mahwah, New Jersey: Lawrence Erlbaum Associates.