

**CBA·NAU**

College of Business  
Administration  
Northern Arizona  
University  
Box 15066  
Flagstaff AZ 86011

# The Impact of Business War Games: Quantifying Training Effectiveness

Working Paper Series 02-20—  
July 2002

**Christopher M. Scherpereel**

Assistant Professor of Management  
College of Business Administration  
Northern Arizona University  
[cms89@mail.cba.nau.edu](mailto:cms89@mail.cba.nau.edu)  
928-523-7831

**CBA·NAU**

College of Business  
Administration  
Northern Arizona  
University  
Box 15066  
Flagstaff AZ 86011

# The Impact of Business War Games: Quantifying Training Effectiveness

Christopher M. Scherpereel

*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 et al. 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 managers 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.<sup>1</sup>

Although some success can be claimed in quantitatively measuring productivity gains from management development programs, 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 (Iaffaldano and Muchinsky 1985; Alliger, Tannenbaum et al. 1997).

Individual’s perceptions or attitude changes are sometimes measured as a proxy indicator of anticipated behavioral changes. Studies by Weigel et al. (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

---

<sup>1</sup> The seminal work in training evaluation can be found in Kirkpatrick (1994). Other studies, focusing on productivity as a training criterion, include Alliger, Tannenbaum et al. (1997) and Shelton and Alliger (1993). For some interesting case study examples see Geber (1995) or Spitzer (1999).

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.

While the quantitative studies measure the wrong thing, productivity changes, the qualitative studies tend to be too subjective to reliably measure the right thing, individual change. People may not perceive changes when they have occurred or may perceive changes when they have not occurred. This study introduces an instrument designed to objectively measure individual perceptions of change and provides a major advancement in the management development training valuation.

The empirical research methodology is tested on the participants in a management-training program that incorporates a business simulation exercise<sup>2</sup>. This program claims to produce the specific outcomes identified as “seeing things differently” and “thinking differently.”<sup>3</sup>

## GENERAL BACKGROUND AND THEORY

### The Simulation

*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).*

Computer business simulations come in a variety of flavors and go by a variety of names. A common use of business simulations is for exploration. These “exploratory” simulations are usually self-contained models of the business and environment that allow the manager to test alternative strategies within the constraints of the model. The managers “use their insights and creativity to think of strategies, then use the simulator to do the (generally complex) math, then revise their strategies based on the model’s feedback; and they go through that process as often as desired” (Reibstein and Chussil 1997, p. 404). This trial and error process helps the managers develop acceptable strategies that can be evaluated and implemented in the actual business.

The second major type of business simulation is called a “rehearsal” simulation,<sup>4</sup> often called a war game or business war game. The rehearsal simulation is used to increase confidence and familiarity with the situation and to find faults with specific strategies.<sup>5</sup> In a business war game, teams of managers are responsible for running different companies in the same industry for a set number of simulated periods (years). A market team (or computer model) is responsible for setting the external market conditions (for example; market growth rates, consumer preferences, total market demand, etc.) and distributing effects to each team based on their decisions. Each team receives performance feedback and relative market position, which they must evaluate and use as a basis for making the next period (year) decisions.

Supporters of 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 and Chussil 1997, p. 409) To bolster

---

<sup>2</sup> The term business simulation exercise, and its synonym business war game, is described in greater detail in the next section. Briefly however, the terminology is used to describe an exercise, developed around a computer simulation of a business, which immerses the participants in a trial and error decision-making environment.

<sup>3</sup> The marketing tag line “see things differently, think differently” is a trademark of PriSim Business War Games, a company that develops and delivers business simulation exercises.

<sup>4</sup> Exploratory and rehearsal simulation are terms used by Reibstein and Chussil (1997).

<sup>5</sup> For those individuals familiar with marketing strategy, the rehearsal simulation is equivalent to test marketing a product. The product is the participant’s strategy and the test market is the rehearsal simulation. The product (strategy) is tested under simulated conditions in a test market (computer-simulation) before it is launched in the larger target market (implemented as part of the company strategy).

this claim, professors David Reibstein and Mark Chussil cite several case studies based on post-exercise interviews and their experience. A condensed version of one of these case studies taken from Day and Reibstein's (Day and Reibstein 1997) book, "Wharton on Dynamic Competitive Strategy" is paraphrased below:

*The vice president was looking for methods to turning around the companies sinking profits. Managers had insisted that the firm's product had become a commodity business, and projected declines in prices meant that the firm anticipated a five-year loss of \$500 million. The managers believed they could reverse this prognosis by cutting costs. Using a business simulation, the managers identified that this strategy would still result in a \$450 million loss. The simulation highlighted that the manager's tendency to compete on price led to the customer realizing the benefits, rather than the firm. This outcome convinced the managers that they needed to do something different. The simulation gave them a tool to test alternatives before risking the implementation in real life. The result according to Day and Reibstein was a differentiated product and profitability (Day and Reibstein 1997, p. 409-410).*

Case studies do seem to support the 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.<sup>6</sup> Even though case studies support these claims, practitioners who have a vested interest in promoting the technique have participated in most of these existing studies.<sup>7</sup> Formal studies that actually attempt to measure the benefits of a simulation exercise are limited and have produced mixed results.<sup>8</sup> This experimental study is the first of a series to look at the actual impact of a business war game exercise intervention on the change that participant's experience in the way they see their business and the way they think about their business.

### **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 have similar responsibilities. These participants are divided into four groups by the executive management.<sup>9</sup>

The intervention is a custom business war game that provides a rehearsal simulation environment for the teams of participants to run the branch office of a service company. The language and dynamic of the war game are designed to simulate the actual environment with which these participants are familiar. The teams of participants are given the mission to develop and execute strategies for the simulated branch office for a period of five years. At the end of this period, the team that performs "the best," as judged by the executive management team, will be crowned the winner<sup>10</sup>.

The environment is competitive. Each team is given a similar branch office to start the war game. They are also given some public information on the environment and competitive landscape. Each year of the war game is interspersed with mini-lectures highlighting the common techniques for analyzing both the internal and external conditions that the branch office is experiencing. Specifically, there are mini-lectures on team dynamics, finance,

---

<sup>6</sup> A number of specific case studies have been published to promote the benefits of business war game exercises, these include: (Gwynne 1995; Hequet 1995; Wilson and Condom 1995; Sherman 1996; Lefebvre 1997; Reibstein and Chussil 1997; Stewart 1997; McCune 1998; McIlvaine 1999).

<sup>7</sup> This includes Reibstein and Chussil (1997), who are active partners in a firm that builds and sells business war games.

<sup>8</sup> To review the literature on this continuing debate see Randel et al. (1992) or more recently Chapman and Sorge (1999). Other important references in the field of measuring business war game effectiveness are (Wolfe 1985; Keys and Wolfe 1990; Anderson and Lawton 1992)

<sup>9</sup> Random team membership is not utilized because the small sample size created concern that several of the individuals, who were deemed "strong" by the executive management, could be placed on the same team. The executive management is in the best position to assure that the final team makeup is homogeneous in terms of individual strengths and weaknesses.

<sup>10</sup> Management selected criteria included a weighting; market-share (targeted), return on investment (profitability), employee satisfaction (turnover), employee utilization, and customer satisfaction.

marketing, and strategy. At the start of each simulated year, teams are debriefed on their relative position and performance. They are then given the new starting positions and directed to breakout rooms where they can formulate decisions for the next simulated year. Participants are completely immersed in the war game for three full days; a minimum of eight hours of formal activity is scheduled for each day.

The custom war game is specifically designed to imitate a branch office operation in a competitive environment. 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 intervention, briefly described above, is the typical package called a business war game or business simulation exercise. In this application it is being used as a development exercise, to improve or enhance participants' performance in their existing positions. It is also commonly used as a training exercise to expedite the development of skills in a new position or industry. These are very different purposes and the reader is warned not to generalize specific results beyond the constraints of this study. More details on the study constraints will be given in a later section when specific parameters are addressed.

### **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.<sup>11</sup> The most common among these reported benefits include:

- Time compression – years become hours
- Decisions without risk – poor performance does not result in a tangible loss
- Promotes creativity – low (no) risk environment promotes experimentation
- Immediate feedback – participants experience quickly the results of their decisions
- Competitive focus – immediate feedback focuses participants on competitive activity
- Identify real information needs – what information is required to make specific decisions
- Cross-functional understanding – generates a common language
- Confidentiality – scenarios can be developed and tested in confidence

Although case study and experience support these benefits,<sup>12</sup> 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 (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 and 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.

One 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. First, the study identifies changes in individual perceptions of decision problems along several key semantic dimensions<sup>13</sup>, verifying that the intervention

---

<sup>11</sup> See (Lefebvre 1997; Reibstein and Chussil 1997; Chapman and Sorge 1999).

<sup>12</sup> The benefits listed are typically designed into the exercise. Thus, they represent the business war game's design criteria and not necessarily the actual benefits gained from participation in such an exercise.

<sup>13</sup> This study is based on the semantic differential technique developed by Osgood et. al. (1957).

has in fact resulted in a significant change. Simultaneously, 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,  $X_n$  represents the decision targets,  $C$  corresponds to the "decision characteristic" scales, and  $A$  denotes the "decision approach" scales):

- H1: Decision target [  $X_n$  ] is characterized as ordinally more/less [  $C$  ] after participation in the business simulation exercise.
- H2: Decision target set [  $X_1, X_2, X_3, X_4, X_5, X_6$  ] is reordered along the [  $C$  ] dimension after participation in the business simulation exercise.
- H3: Decision problem [  $X_n$  ] is approached ordinally more/less [  $A$  ] after participation in the business simulation exercise.
- H4: Decision target set [  $X_1, X_2, X_3, X_4, X_5, X_6$  ] 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. Generating this list of specific a-priori hypotheses is the first major task in preparation for the experimental field study. Having generated this listing, the objective can be clearly 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 the eight 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 and 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. Campbell and Stanley (1963) identify three experimental design groupings based on the design's ability to control for parameters: 1) pre-experiments, 2) true experiments, and 3) quasi-experiments.<sup>14</sup>

This study uses a design from the most limiting of this group, the 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.

---

<sup>14</sup> For complete details on experimental designs, see Campbell and Stanley (1963).

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<sup>15</sup> 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

### Target Decision Selection

The target decisions 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<sup>16</sup>.

Table 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.

BUILDING Competencies SETTING planned sales volume <div style="border: 1px solid black; padding: 2px;">ESTABLISHING Strategy</div> <div style="border: 1px solid black; padding: 2px;">HIRING a Loan Officer</div> <div style="border: 1px solid black; padding: 2px;">ALLOCATING Loan Officer Time</div> ALLOCATING Area Sales Manager (ASM) Time <div style="border: 1px solid black; padding: 2px;">TRAINING Loan Officers</div>	FORECASTING Sales MANAGING Employee Turnover <div style="border: 1px solid black; padding: 2px;">TARGETING Customers</div> SELECTING Measures of Business Success SELECTING Tactical Focus RESPONDING to Competitor Actions <div style="border: 1px solid black; padding: 2px;">SOURCING Leads</div>
---	--

**Table 1: Target Decision Concepts**

The six target, boxed, decision concepts selected from Table 1 are abbreviated as follows: Strategy (establishing strategy), Hiring (hiring a loan officer), Time Block or Time Blocking (allocating loan officer time),

<sup>15</sup> The term “remote location” is used in this study to describe a location away from the interruptions of daily work demands. In this particular case, all participants resided in a hotel/conference facility for the duration of the exercise.

<sup>16</sup> For a more complete discussion of the constraints placed on the design and development of a business war game see (Goosen 2001)

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 of eight bipolar semantic differential scales are developed as the measurement instruments for the “decision characteristic” and “decision approach” concepts in Figure 1. These scales are applied across six specific decision targets for a total of ninety-six measurements per respondent. The 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 “decision characteristic” and “decision approach” concepts’ meaning.

TRAINING Loan Officers		TRAINING Loan Officers	
Determining how much effort should be spent training loan officers is a(n) _____ decision for the branch office?		Determining how much effort should be spent training loan officers requires a(n) _____ approach?	
Simple _____	Complex _____	Textbook _____	Gut (intuition) _____
Long term _____	Short term _____	Quick _____	Slow _____
Reversible _____	Irreversible _____	Passive (search) _____	Active (search) _____
Unimportant _____	Important _____	Methodical _____	Haphazard _____
High risk _____	Low risk _____	Team _____	Individual _____
Constant (static) _____	Changing (dynamic) _____	Risk avoiding _____	Risk taking _____
Big _____	Small _____	Big picture _____	Detailed _____
Clear _____	Ambiguous _____	Planned _____	Unplanned _____

Figure 1: Experimental Field Study – Question/Layout

### Layout

The decision dimension questionnaire begins with a page of detailed instructions, describing the correct marking of the instrument. The instructions are presented in three sections; the general information, specific example and important notes;<sup>17</sup> and the landscape layout is used to conform to the space requirements of the dual column question format. 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 1.

### Target Decision Order

Six target decisions are included in the decision dimension questionnaire. The placement of each target is consistent with the suggestion to place easier decisions earlier in the questionnaire (Emory 1985, p. 222). Therefore, the first target is selected because it is the most familiar among the participants. The decision familiarity decreases as the respondent moves from decision target number one through decision target number four. The final decisions, five and six, return to a moderate level of familiarity.

It is worth noting that the first three decision targets can be categorized as tactical decisions, having a more immediate impact on the business, while the last three decision-targets are primarily strategic. This order allows the respondent to answer the easier tactical decisions before having to switch to the more difficult strategic decisions. The fourth decision target, “establishing strategy,” is designed to help the respondent switch from a tactical mindset to a strategic mindset as he/she completes the questionnaire.

### Scale Order and Polarity

The eight semantic differential scales, for each of the two decision concepts, are pseudo-randomly ordered. A heuristic is applied that attempts to separate scales that are intuitively similar, or were previously found to measure the same dimension. If the scales cannot be separated, then an attempt is made to reverse the polarity of one of the scales. Reversal of scale polarity is also done to select scales, when the interpretation of the scale would not be significantly

<sup>17</sup> For details on the instruction design, see the pilot test design details.



impacted. For example, the polarity of the scale “small-big” is reversed to “big-small” without significantly impacting the scale’s difficulty, since the opposite pole is obvious given either term. In contrast, the scale simple-complex is intuitively more appealing than the reversed scale complex-simple. Therefore, the scale simple-complex is not considered for reversal. The goal of the heuristic is to keep a term that is most “familiar”<sup>18</sup> to the respondent on the left-hand pole of the scale. This ordering heuristic procedure encourages the respondent to read the polar terms carefully before marking the scale.

Scale order and polarity is maintained from decision target to decision target. Although increasing the repetitiveness of the questionnaire, this feature speeds the marking process. It also provides some assurance that each decision target is receiving similar consideration by the respondent.

### **Additional Design Features**

Since this study is designed to gather data on individual change, a space for the respondent’s name is included on the first page of the questionnaire. The respondent’s name is then used to match his/her pretest with his/her posttest questionnaire, which is critical for analysis. The posttest questionnaire is identical to the pretest questionnaire except for an indication that the questionnaire is a follow-up. The posttest questionnaire presents the concepts and the associated semantic differential scales in exactly the same format as the pretest questionnaire. This avoids some of the instrumentation bias<sup>19</sup> concerns that might be introduced by altering the design.

A cover letter is included with each questionnaire to encourage response. On the initial questionnaire, the cover letter is from the company’s director of training, introducing the research and its importance to the business. The cover letter attached to the posttest questionnaire is a letter from the researcher, thanking the respondents.

### **Administration**

The experimental field study is administered in two parts. The program facilitator distributes the pretest questionnaire at the beginning of the first day, of a three-day business war game exercise. Instructions are given to return the completed questionnaires at the first program break. Fifteen minutes is explicitly allocated for completing the questionnaire, with additional time available during the participant’s break. Both the program director and the firm’s executive management verbally highlight the importance of the study, which is reinforced in the cover letter. The questionnaires are collected and the inclusion of the respondent’s name is verified visually.

The posttest questionnaire is distributed during the final session of the third day. Again, fifteen minutes are allocated for completing the questionnaire, with additional time available at the session’s end. Instructions are given to return the questionnaires before departing, however, in the event that it cannot be completed, an alternative collection method is offered.<sup>20</sup>

### **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<sup>21</sup> 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”

---

<sup>18</sup> Scale reversal is restricted to scales that have a common polar opposite. This implies that the term is “familiar” to the respondent. In other words, the scale’s left-hand term is most likely to be in the respondent’s common language set and there is little confusion about its meaning.

<sup>19</sup> Instrumentation bias is a threat to the internal validity of the experimental design. This bias is the result of changes between the pretest and posttest measurements, which arise from changes in the instrument design.

<sup>20</sup> In this study, all participants attending the final session were able to complete the posttest questionnaire prior to departure. However, there was one participant, who for emergency reasons, needed to leave just prior to the final session and did not receive the posttest questionnaire. Because of the small sample size, it was decided that the inclusion of this participant’s posttest responses was important. Therefore, the questionnaire was immediately sent electronically to the missing respondent. This questionnaire was electronically returned within a week and carefully inspected for abnormalities. None were detected, so these data were added to the full data set without further note.

<sup>21</sup> There are six target decisions and two target concepts each having eight semantic scales, for a total of  $6 \times 2 \times 8 = 96$  possible combinations.

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 Table 2.

Independent Variable (Decision Target) X	Dependent Variable C	Direction of movement* <, > <+>, <0>	Raison d'être
<b>Characteristic</b>			
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
<b>Approach</b>			
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
<b>Notes</b>		<b>* Key</b>	
* "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

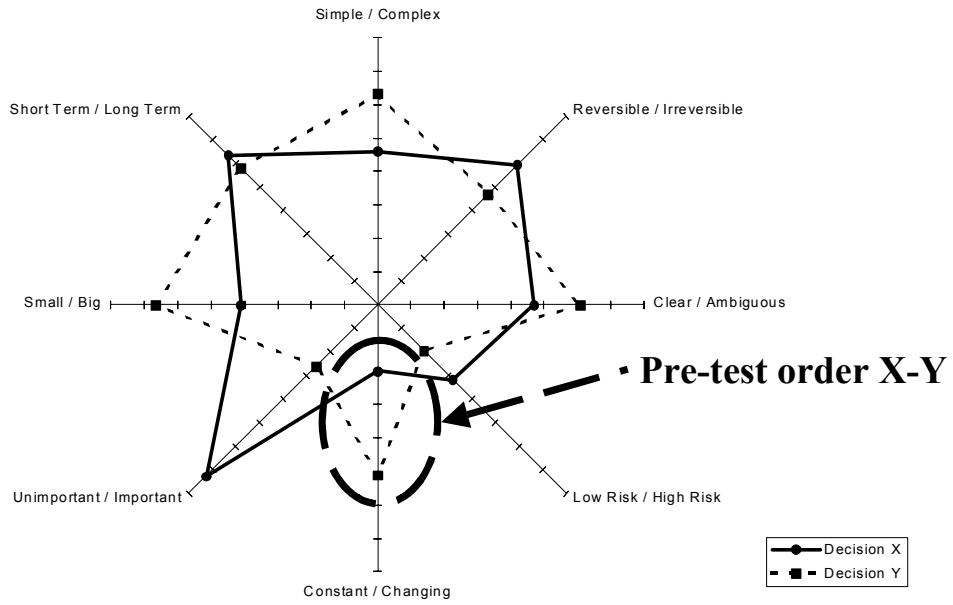
**Table 2: A-priori Hypotheses for “Targeted” Changes**

Only sixteen hypotheses can be generated from H2 and H4.<sup>22</sup> In this case, the pre-test data will be used as the basis for the a-priori ordering<sup>23</sup> and it will be compared to the post-test data to validate the hypotheses acceptance. Figure 2 and Figure 3 are generated using fictitious data to help visualize the change in order implied by these hypotheses. However, to simplify the presentation only two decision-targets are illustrated. The actual data, having six decision-targets, will be presented in a similar format in the “results” section.

<sup>22</sup> There are six target decisions ordinally ranked by the pre-test data along the eight semantic scales for both the “decision characteristic” and “decision approach” concepts. This results in total number of combinations of 1x8x2 =16. It might be possible to generate a more carefully conceived a-priori hypothesis where each decision target is tested for change against every other decision target. However, without prior art indicating a-priori what these changes might be, the task would be ominous. An effort like this would require the research to specify all combinations of the six decision-targets along the eight semantic scales for both concepts, for a total of 6x6x8x2=576 a-priori hypotheses.

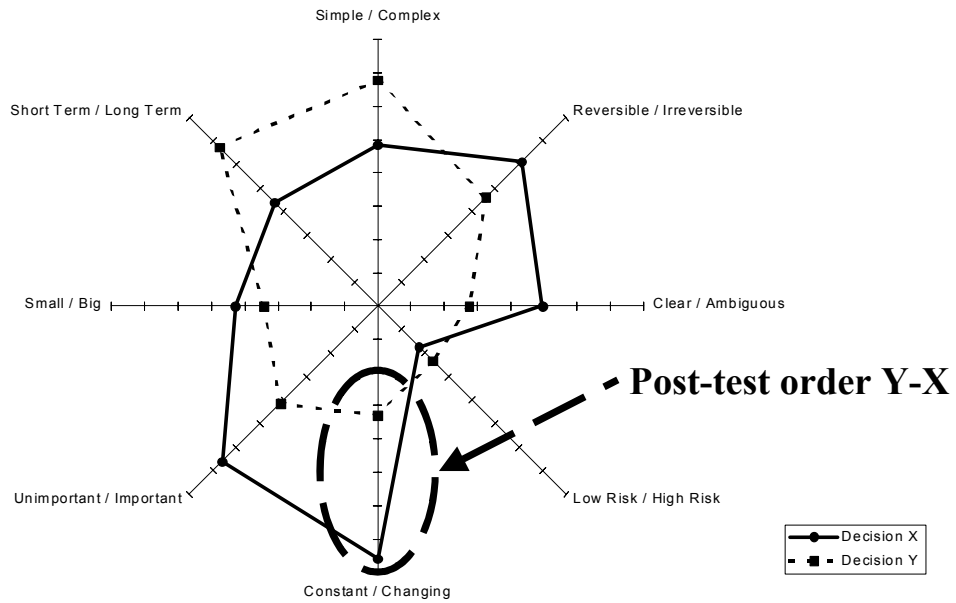
<sup>23</sup> During the analysis of the “pre-test ordering data,” tests will be conducted to validate that the ordering has some level of confidence.

**Pre-test Decision Target Order (For Illustration Purposes ONLY)**



**Figure 2: Illustrative Pre-test Decision Target Order**

**Post-test Decision Target Order (For Illustration Purposes ONLY)**



**Figure 3: Illustrative Post-test Decision Target Order**

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.<sup>24</sup> 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.

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.<sup>25</sup> 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.

### LIMITATIONS

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.

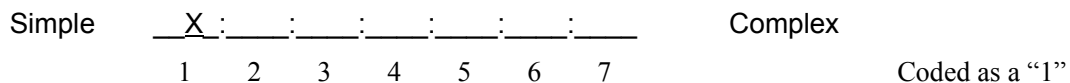
### DATA PREPARATION

The next several sections present the data coding and editing procedures. Focus is placed on communicating a consistent understanding of the data set.

#### Data Coding

Immediately after the questionnaires are collected, a preliminary verification process is administered. This process includes the confirmation that the questionnaire is complete, containing at least the three decision concept pages (the questionnaire is considered complete without the cover page of instructions), and a check that the name included is the same as that recorded on the official participant list. The checked participant list is then examined to verify that each participant has returned a completed pretest and posttest questionnaire.

A coding of the semantic concept scales follows completion of this preliminary verification. The semantic differential scales are prefaced with the letter "Q" and numbered sequentially from one (1) to forty-eight (48), for the "decision characteristic" concept, and forty-nine (49) to ninety-six (96), for the "decision approach" concept. During the initial coding process, data from the pretest and posttest questionnaires are maintained in separate data files and number-coded identically. To reduce coding errors and ease data entry, the semantic differential scales are then scored from "1," representing the left-most pole, to "7," representing the right-most pole. An example of the coding for the simple-complex semantic scale is illustrated in Figure 4. Any missing data was coded with a "\*" and excluded from the data set.



**Figure 4: Experimental Field Study Questionnaire Coding**

<sup>24</sup> See (Csikszentmihalyi 1996; Sherman and Schultz 1998; Evans and Wurster 2000).

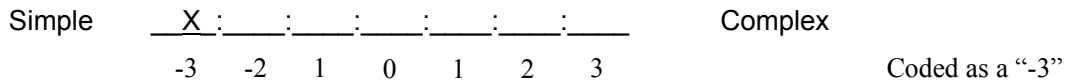
<sup>25</sup> See (Zsombok 1997; Klein 1998).

After completing the initial coding and rechecking for errors, the data is transformed to simplify the analysis. First, each semantic scale is uniquely identified using the key reproduced in Table 3. For example, scale “Q1” from the pretest questionnaire is re-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. This re-labeling eases the concern with making errors in locating specific variables in the large data set. The pretest and posttest data sets can now be combined and analyzed without confusion.

Decision Targets		Label	CODE	Characteristic Scales		CODE
HIRING a Loan Officer		<i>Hiring</i>	<b>H</b>	Simple - Complex		<b>S/C</b>
TRAINING Loan Officers		<i>Training</i>	<b>TR</b>	Short Term-Long Term		<b>S/L</b>
ALLOCATING Loan Officer Time		<i>Time Blocking</i>	<b>TB</b>	Reversible - Irreversible		<b>R/I</b>
ESTABLISHING Strategy		<i>Strategy</i>	<b>ST</b>	Unimportant-Important		<b>U/I</b>
SOURCING Leads		<i>Sourcing</i>	<b>SO</b>	Low Risk-High risk		<b>L/H</b>
TARGETING Customers		<i>Targeting</i>	<b>TA</b>	Constant-Changing		<b>C/C</b>
				Small-Big		<b>S/B</b>
				Clear-Ambiguous		<b>C/A</b>
Questionnaires		CODE		Approach Scales		CODE
Pretest		-		Textbook-Gut		<b>T/G</b>
Posttest		+		Quick-Slow		<b>Q/S</b>
Ordinal/Interval Label	Score (Left or Right)			Active-Passive		<b>A/P</b>
extremely	<b>-3 or +3</b>			Methodical-Haphazard		<b>M/H</b>
quite	<b>-2 or +2</b>			Individual-Team		<b>I/T</b>
slightly	<b>-1 or +1</b>			Risk Avoiding-Risk Taking		<b>R/R</b>
equally	<b>0</b>			Detailed-Big Picture		<b>D/B</b>
				Planned Unplanned		<b>P/U</b>

**Table 3: Experimental Field Study Questionnaire Coding and Labeling Key**

Having re-labeled the data sets, the polarities of the concept scales that had been intentionally reversed during the design process are restored. As suggested by Osgood et al. (1957), the coding is transformed to a “-3” to “+3” scoring, with the central point on the scale designated as “0” to represent neutrality. Under this transformation, 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 5 illustrates this coding transformation, which can be easily compared to the initial coding in Figure 4.



**Figure 5: Experimental Field Study Questionnaire Transformed Coding**

### Data Editing

In this study, data editing is relatively simple. 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,<sup>26</sup> or a long sequence of scales is marked at the same level.<sup>27</sup> 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.

<sup>26</sup> For example, a check is made to verify that not all the simple-complex scales are rated at the same level for all decision targets.

<sup>27</sup> For example, if all the scales under a particular decision target are rated at the same level.

## DATA ANALYSIS AND FINDINGS

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.

The raw results are reformulated using the tools developed to provide a visualization of the change resulting from participation in the business war game intervention. These visualizations provide the tool needed to quickly evaluate the merits of hypothesis H2 and H4, as well as provide additional support for H1 and H3.

In this section, a number of statistical tools will be used to visualize and parameterize 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. A matched pairs t-test is performed to assess significance, and these data are presented in a series of spider charts to visually present the results. For brevity, the summary statistics and data presented in this section will be restricted to the twelve measurements identified in Table 2, which are called the “targeted twelve.”

### Pretest and Posttest Descriptive Statistics Summary

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 4 and Table 5 provide a results summary, limited to the “targeted twelve,” for the pretest and posttest data respectively.

Scale	Count	Minimum	Maximum	Mean	Ordinal Descriptor	Median	Mode	Standard Deviation	Skewness	Kurtosis	Normal Probability	Rootogram
<b>"Decision Characteristic" Concept</b>												
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
<b>"Decision Approach" Concept</b>												
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 4: “Targeted Twelve” Pretest Summary Statistics**

Scale	Count	Minimum	Maximum	Mean	Ordinal Descriptor	Median	Mode	Standard Deviation	Skewness	Kurtosis	Normal Probability	Rootogram
<b>"Decision Characteristic" Concept</b>												
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
<b>"Decision Approach" Concept</b>												
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

**Table 5: “Targeted Twelve” Posttest Summary Statistics**

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 “targeted twelve” are presented in Table 6.

Scale	Count	Minimum	Maximum	Mean	Ordinal Descriptor (Mean)	Median	Mode	Standard Deviation	Skewness	Kurtosis	Normal Probability Plot	Rootogram
<b>"Decision Characteristic" Concept</b>												
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
<b>"Decision Approach" Concept</b>												
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

**Table 6: “Targeted Twelve” Gap (Posttest – Pretest) Summary Statistics**

### Analysis of Change (Aggregated)

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, or the non-parametric Wilcoxon signed-rank test on the gap data file.

The matched t-test of the means,<sup>28</sup> 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 ≠ pretest mean – posttest mean ≠ 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), versus the alternative hypothesis H1: mean difference < 0 (or > 0). These results are tabulated for the “targeted twelve” in Table 7. 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.

<sup>28</sup> The important assumption is that the mean difference follows a normal distribution, an assumption not verified in this study.

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

**Table 7: "Targeted Twelve" Hypothesis Results (matched paired t-test)**

Alternatively, the one-sample Wilcoxon signed-rank test<sup>29</sup> 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 ≠ hypothesized median gap ≠ 0. These results are tabulated for the "targeted twelve" in Table 8. 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.

<sup>29</sup> An assumption for the one-sample Wilcoxon test is that the data are a random sample from a continuous, symmetric population, an assumption that is difficult to claim for the discrete data set used in this analysis. When the population is normally distributed, this test is slightly less powerful (the confidence interval is wider, on the average) than the t-test. It may be considerably more powerful (the confidence interval is narrower, on the average) for other populations.



**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 (>)
<b>"Decision Characteristic" Concept</b>							
H S/B	21	10	32.5	0.0	0.65	0.71	0.32 *
TR S/B	21	11	54	-0.5	<u>0.07</u>	0.97	<u>0.03</u> *
TB S/L	21	14	84.5	-0.5	<u>0.05</u>	0.98	<u>0.02</u> *
ST C/A	21	14	26.5	0.5	<u>0.11</u>	<u>0.06</u> *	0.95
SO C/C	21	10	5	0.5	<u>0.03</u>	<u>0.01</u> *	0.99
TA C/C	21	10	13	0.5	<u>0.15</u>	<u>0.08</u> *	0.94
<b>"Decision Approach" Concept</b>							
H D/B	21	14	78.5	-1.0	<u>0.11</u>	0.95	<u>0.06</u> *
TR D/B	21	9	41	-1.0	<u>0.03</u>	0.99	<u>0.02</u> *
TB D/B	21	17	118	-1.5	<u>0.05</u>	0.98	<u>0.03</u> *
ST T/G	21	16	44	0.5	0.22	<u>0.11</u> *	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	<u>0.16</u>	0.93	<u>0.08</u> *

**Table 8: "Targeted Twelve" Hypothesis Results (Wilcoxon signed-rank test)**

**Analysis of Change (Individual Cases)**

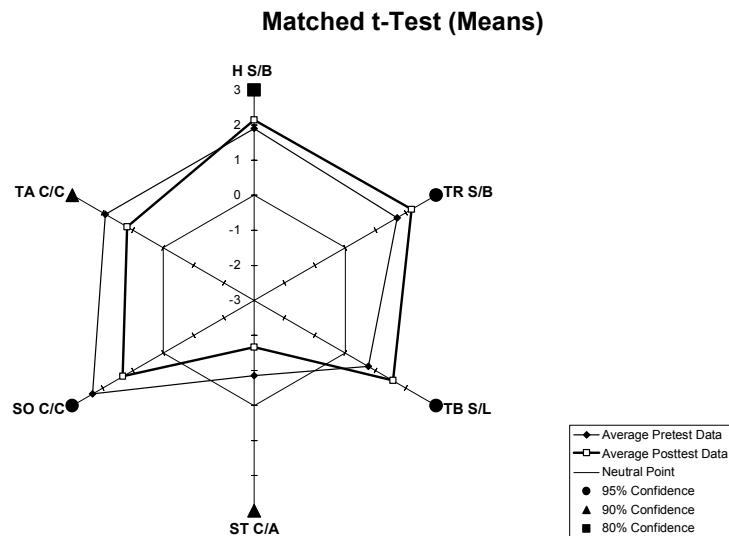
The gap data can also be used to identify specific changes in individuals. A non-statistical heuristic measure is simply the frequency, or count of values exceeding certain changes. Snider and Osgood (1969) states that "the evidence shows that for individual subjects a shift of more than two scale units 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 exceed two scale units as "significant" when analyzing individual cases. Table 9 contains these data for the "targeted twelve" semantic differential scales.

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)
<b>"Decision Characteristic" Concept</b>					
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%
<b>"Decision Approach" Concept</b>					
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%

**Table 9: "Targeted Twelve" Individual Cases – Summary of Change**

### Spider Chart Analysis

An effective way to visualize these multi-scale changes is to use spider charts. The spider charts equivalent to tabulating the pretest and posttest data, and the hypothesis test results, are illustrated in the following four figures: Figure 6, Figure 7, Figure 8, and Figure 9. This format allows a quick visual comparison to be made between the matched pair t-test and the Wilcoxon signed-rank test results. The complete results are presented using an alternative format in a later section.



**Figure 6: "Decision Characteristic" Concept – Multiple Decision Targets (matched pair t-test)**

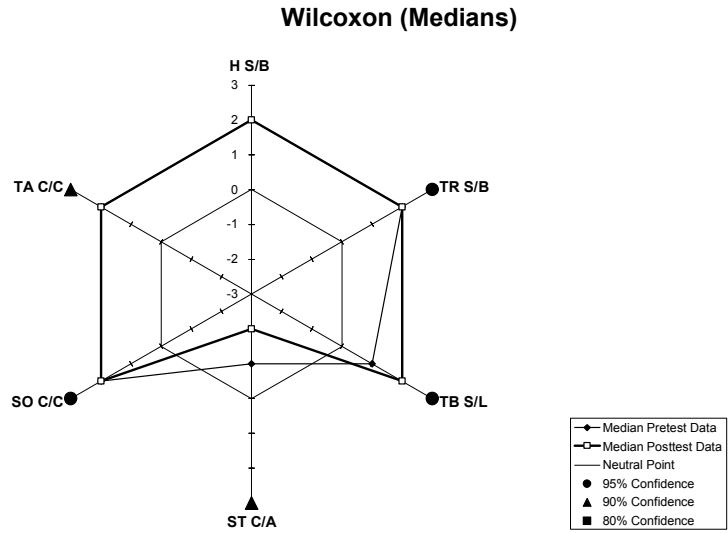


Figure 7: “Decision Characteristic” Concept – Multiple Decision Targets (Wilcoxon signed-rank test)

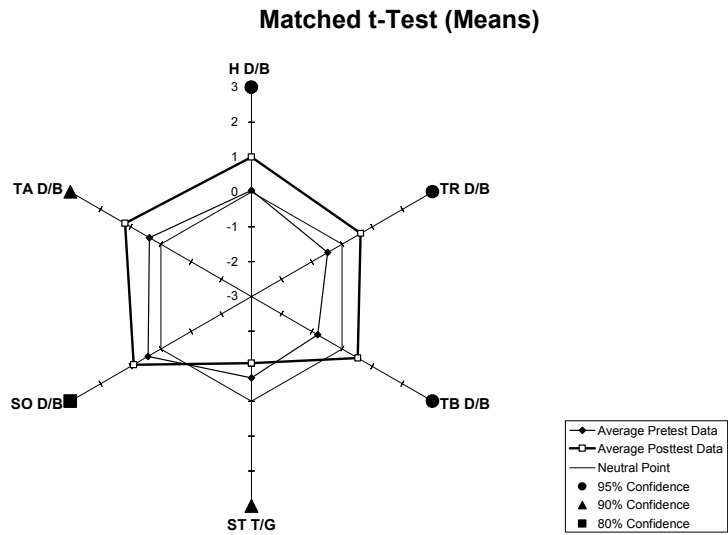
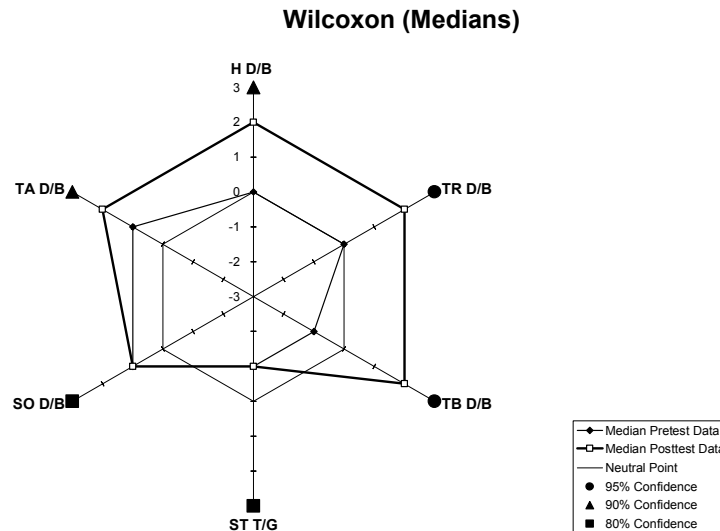


Figure 8: “Decision Approach” Concept – Multiple Decision Targets (matched pair t-test)



**Figure 9: “Decision Approach” Concept – Multiple Decision Targets (Wilcoxon signed-rank test)**

### Test for Change

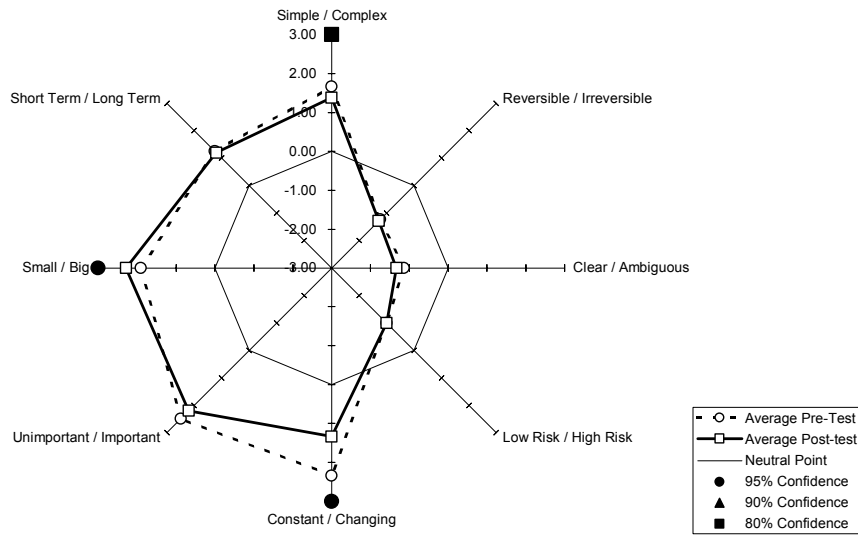
The minimization processes, used for both the cluster and the factor analysis techniques, result in a convenient semantic space for displaying the multiple dimensions of each decision target. In the previous analysis, focus was placed on the “targeted twelve” measures. Now the focus shifts toward the changes in the participant’s perceptions of the actual six decision targets while considering all measures. First, a profiling methodology is introduced, with examples of selected decision targets. Then the decision target rank order hypotheses, H2 and H4, are directly addressed using this methodology.

### *Profiling Methodology - Change along Specific Dimensions*

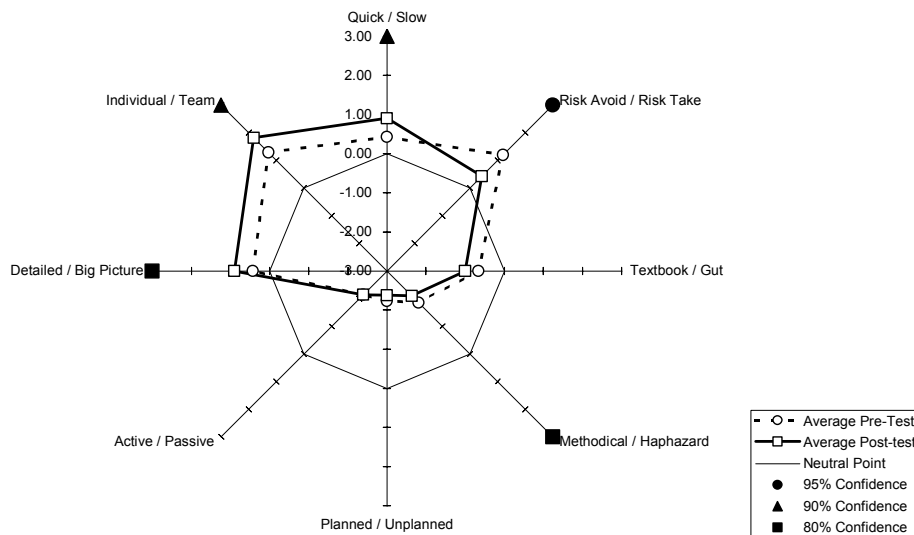
This visual profiling provides a convenient method for identifying the individual participant’s perception of change, and when aggregated the perceived group change. It is useful to examine the important “sourcing” decision target in detail.<sup>30</sup>

The semantic space defined in the level-two analyses is used as the basis for profiling the aggregate group data for each of the decision targets. In this profiling, the statistical measure describing the data is placed along the axis and a line is drawn connecting these data. If the change is identified as being significant, achieving the designated level of confidence, then an indicator is placed at the end of the scale to highlight the level of confidence achieved. The aggregated results for the “sourcing” decision target are displayed in Figure 10 for the “decision characteristic” concept and Figure 11 for the “decision approach” concept. The effect noted in these figures is that the profile appears to be stretched like a rubber band in a general direction rather than exclusively along individual measures.

<sup>30</sup> These illustrations use the matched pair t-test and the two-tailed hypothesis test to assess the confidence levels. This is a more conservative depiction than might be derived from using the different single-tail hypotheses for each dimension, or using Wilcoxon signed-rank test of the median.

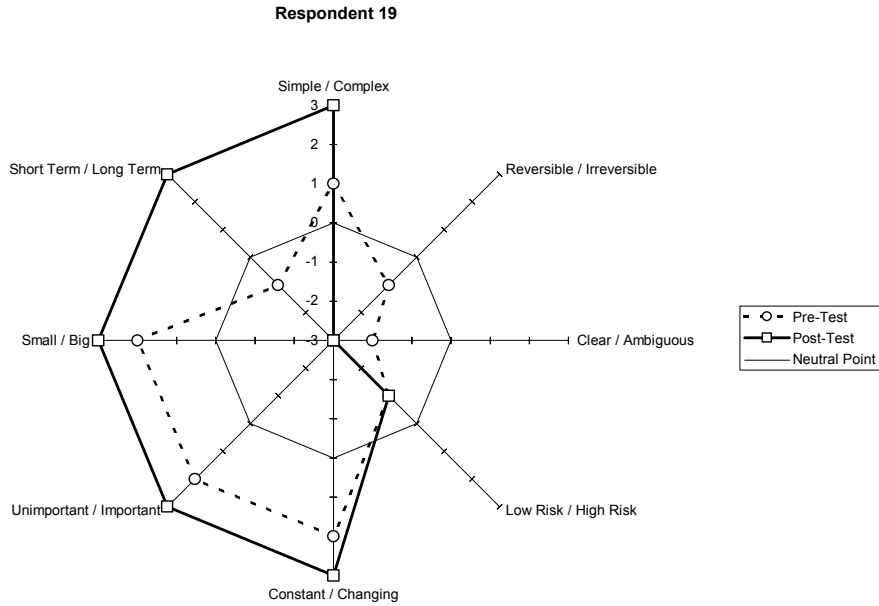


**Figure 10: Profiling the Sourcing “Decision Characteristic” Concept – Group**

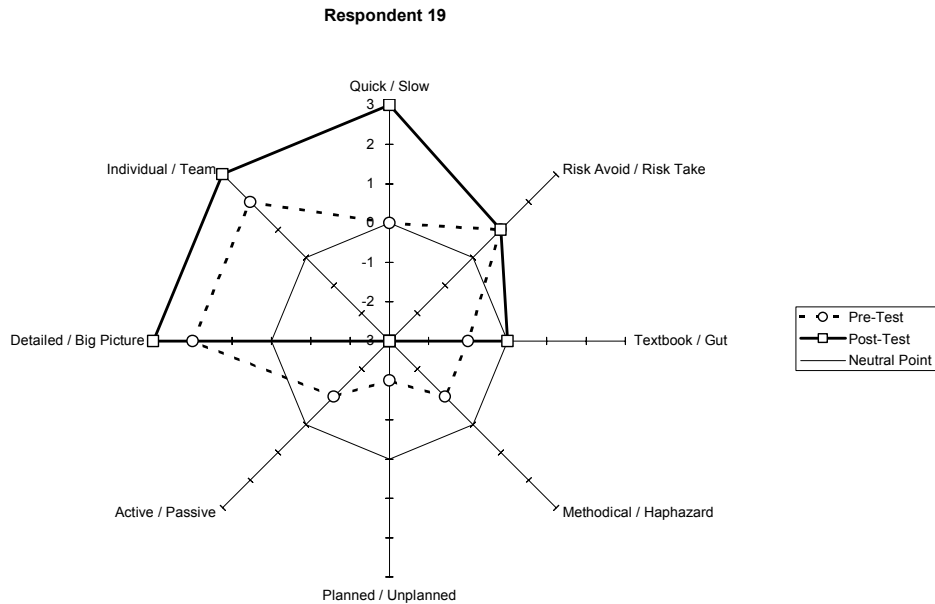


**Figure 11: Profiling the Sourcing “Decision Approach” Concept – Group**

An individual can be profiled in the same manner, using the absolute measures taken directly from the questionnaire instruments. This is useful, from a diagnostic perspective, to identify individuals that have perceived changes in the decision concepts not identified in the aggregated results. The “sourcing” profiles for a particular individual are generated in Figure 12 and Figure 13, for the “decision characteristic” and “decision approach” concepts, respectively.



**Figure 12: Profiling the Sourcing "Decision Characteristic" Concept - Individual**



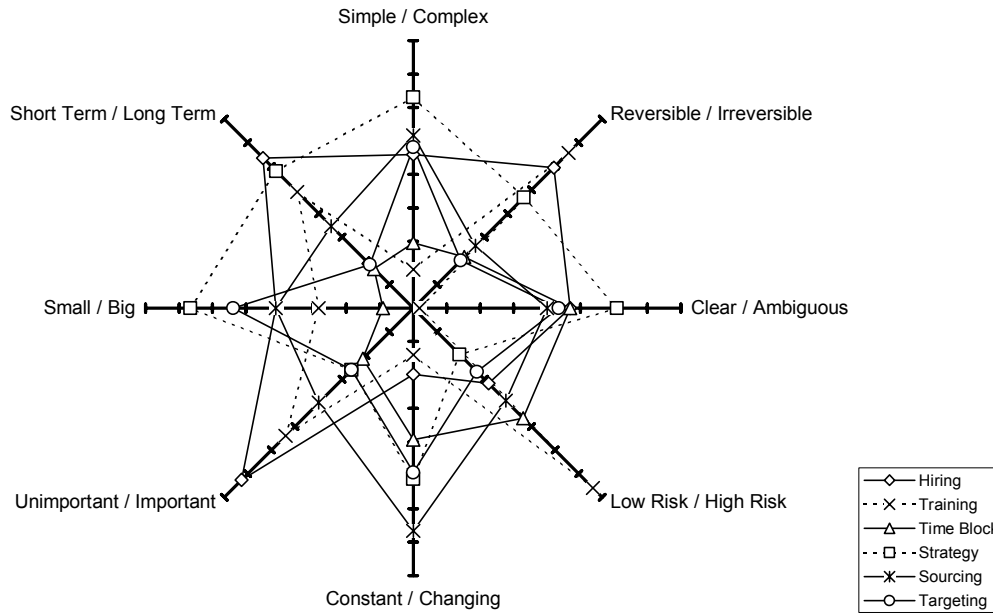
**Figure 13: Profiling the Sourcing "Decision Approach" Concept - Individual**

*Change in Perceptual Ordering*

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 used for profiling the semantic space. 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.

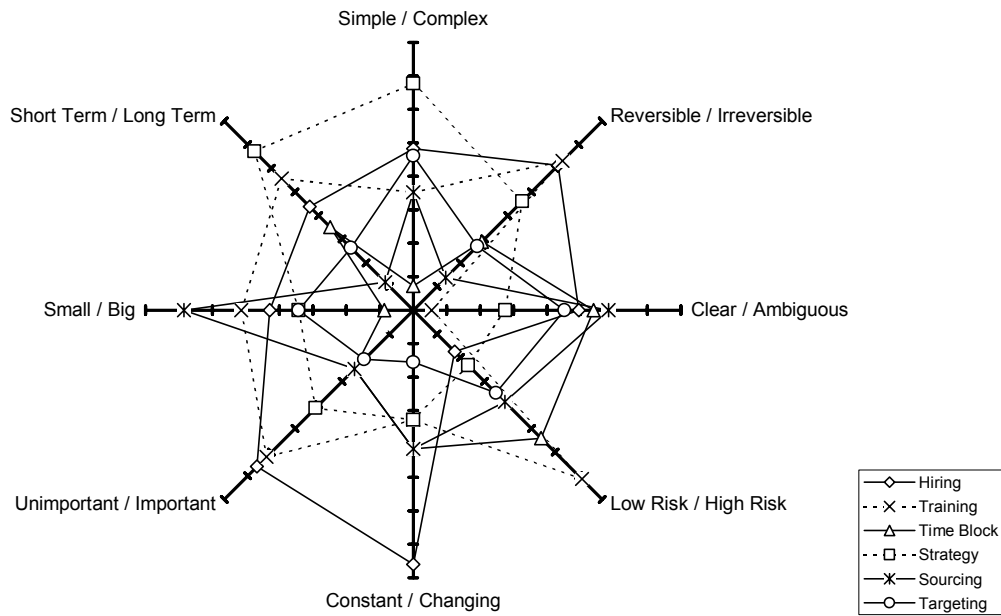
Statistical tools, which require the assumption of interval data, are again used in the assessment. The mean and t-test statistics are chosen to evaluate the magnitude of the change and whether the change is significant. To visually summarize these findings, the data are normalized and illustrated on the semantic space spider charts. Figure

14 and Figure 15 can be visually compared to determine if any ordering change in the decision targets for the “decision characteristic” concept has occurred.<sup>31</sup> 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 on the graphics.



**Figure 14: Target “Decision Characteristic” Pretest Ordering**

<sup>31</sup> The scale values have been removed to avoid any confusion that the magnitude (or the distance from the neutral point on the scale) indicated on the pretest graphic can be compared to the magnitude on the posttest graphic. This comparison can only be made with non-normalized data.

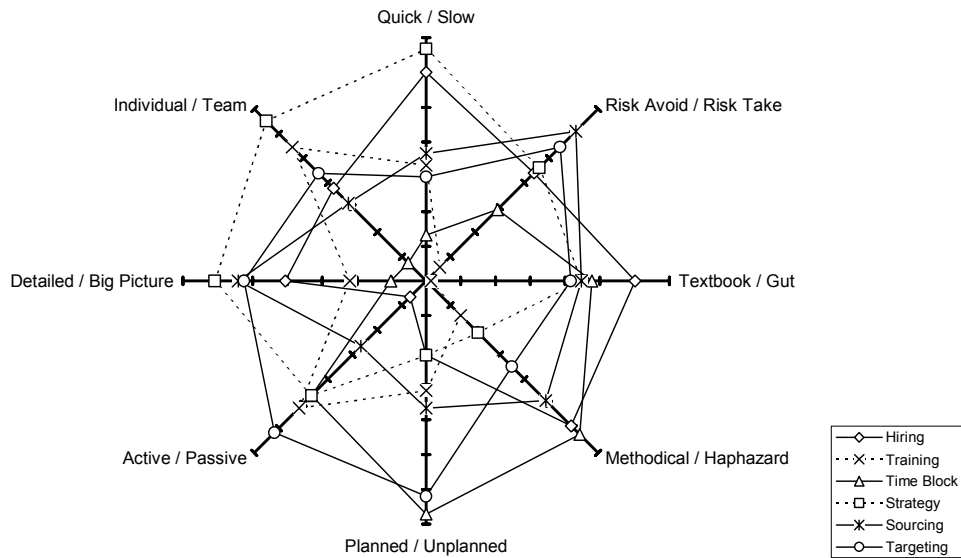


**Figure 15: Target “Decision Characteristic” Posttest Ordering**

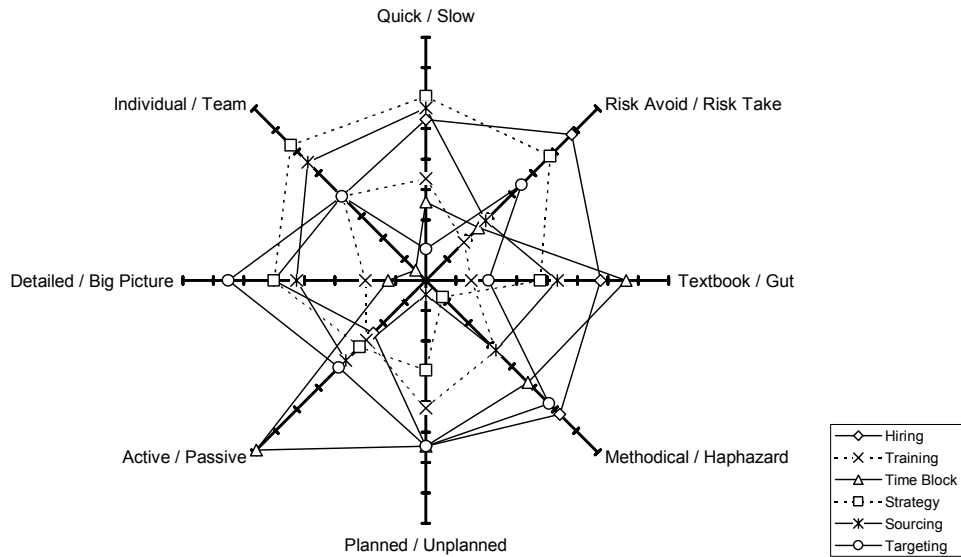
The same procedure is used to assess ordering changes in the “decision approach” concept. Again, the data are normalized, so the only valid comparison is in the decision target order within each of the two figures, Figure 16 and Figure 17. Tests are conducted to determine significance,<sup>32</sup> and a summary is provided in the discussion of these results.

<sup>32</sup> Statistical t-tests for all order combinations are not included in this paper.





**Figure 16: Target "Decision Approach" Pretest Ordering**



**Figure 17: Target "Decision Approach" Posttest Ordering**

### DISCUSSION

The matched pair t-test confirms ten of the "targeted twelve" hypotheses, at the 90% confidence level, while the Wilcoxon signed-rank test confirms nine. A total of 48 a-priori hypotheses of the form H1 were stated, 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.<sup>33</sup> 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.<sup>34</sup>

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.

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 comparing Figure 14 with Figure 15 and Figure 16 with Figure 17, the data are not sufficient to conclude that the re-orderings are significant. Table 10 and Table 11 convey these results for both the “decision characteristic” and “decision approach” concepts, respectively. 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.

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

**Table 10: “Decision Characteristic” Concept – Decision Target Orders**

<sup>33</sup> The Wilcoxon summed rank test results confirm 12.5% with at least 95% confidence, 25% with at least 90% confidence, and 35.4% confirmed with at least 80% confidence.

<sup>34</sup> The Wilcoxon summed rank test results confirm 10.4% with at least 95% confidence, 18.8% with at least 90% confidence, and 39.6% confirmed with at least 80% confidence.

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

**Table 11: “Decision Approach” Concept – Decision Target Orders**

### 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 indicator 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 principle advantage of this new methodology is that it allows an immediate visual 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.

## CITED REFERENCES

- 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., S. I. Tannenbaum, et al. (1997). "A Meta-analysis of the Relations among Training Criteria." Personnel Psychology **50**: 341-358.
- Anderson, P. H. and L. Lawton (1992). "A Survey of Methods used for Evaluating Student Performance on Business Simulations." Simulation and Games **23**(4): 490-498.
- Campbell, D. T. and J. C. Stanley (1963). Experimental and Quasi-Experimental Designs for Research. Chicago, Illinois, Rand McNally.
- Chapman, K. J. and C. L. Sorge (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., and D. J. Reibstein (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. and T. S. Wurster (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.
- Filipeczak, B., M. Picard, et al. (1998). "Measuring Training's Contribution to Intellectual Capital." Training and Development **35**: 14.
- Geber, B. (1995). "Does your Training make a Difference? Prove It!" Does your Training make a Difference? Prove It! **32**: 27-34.
- 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. and P. M. Muchinsky (1985). "Job Satisfaction and Performance." Psychological Bulletin **97**: 251-273.
- Keys, B. and J. Wolfe (1990). "The Role of Management Games and Simulations in Education and Research." Journal of Management **16**: 307-336.
- Kirkpatrick, D. J. and . (1994). Evaluating Training Programs: The Four Levels. New York, Berrett-Koehler.
- Klein, G. (1993). A Recognition-Primed Decision (RPD) Model of Rapid Decision Making. Decision Making in Action: Models and Methods. G. Klein, J. Orasanu, R. Calderwood and C. E. Zsombok. 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.
- 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., G. J. Suci, et al. (1957). The Measurement of Meaning. Urbana, Illinois, The University of Illinois Press.
- Randel, J. M., B. A. Morris, et al. (1992). "The Effectiveness of Games for Educational Purposes: A Review of Recent Research." Simulation and Games **23**(3): 221-276.
- Reibstein, D. J. and M. J. Chussil (1997). Putting the Lesson Before the Test: Using Simulation to Analyze and Develop Competitive Strategies. Wharton on Dynamic Competitive Strategy. G. S. Day and D. J. Reibstein. New York, John Wiley and Sons, Inc.: 395-423.
- 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: 58.
- Shelton, S. and G. M. Alliger (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. and R. Schultz (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. and C. E. Osgood, Eds. (1969). Semantic Differential Technique: A Sourcebook. Chicago, Illinois, Aldine Publishing Company.
- Spitzer, D. R. (1999). "Embracing Evaluation." Training **36**(6): 42-47.
- Stewart, T. A. (1997). "The Dance Steps get Trickier all the Time." Fortune **135**: 157-158.
- Weigel, R. H., D. T. A. Vernon, et al. (1974). "Specificity of the Attitude as a Determinant of Attitude-behavior Congruence." Journal of Personality and Social Psychology **30**: 724-728.
- Wilson, J. R. and P. Condom (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.
- Zsombok, C. E. (1997). Naturalistic Decision Making: Where Are We Now? Naturalistic Decision Making: Where Are We Now? C. E. Zsombok and G. Klein. Mahwah, New Jersey, Lawrence Erlbaum Associates: 3-16.