Changing mental models: Business simulation exercises
Christopher M. Scherpereel
Simulation Gaming 2005; 36; 388
DOI: 10.1177/1046878104270005

The online version of this article can be found at:
http://sag.sagepub.com/cgi/content/abstract/36/3/388
Changing mental models:
Business simulation exercises

Christopher M. Scherpereel
Northern Arizona University

The effectiveness of business simulation exercises in uncovering the interconnectedness of business decisions has been explored in the literature. Little research has been done, however, to quantify empirically how simply identifying connections can change the decision maker’s perceptions and actions. This research exposes the process by which business simulations bring about these changes. Using the semantic differential technique and a one-group pretest-posttest design, tests are conducted to assess perceptual changes in problem characterization and approach induced by participation in a business simulation exercise. The results presented provide quantifiable justification for the use of business simulation exercises to induce targeted change in a decision maker’s mental models. In addition, business simulation designers are offered a powerful feedback tool that allows them to customize decision parameters and confirm the desired participant impact has been achieved.

KEYWORDS: business simulation; decision making; empirical research; measurement; mental models; perceptions; semantic differential; simulation design.

Business simulation exercises are designed to give participants an opportunity to rehearse strategies and test mental paradigms. Case studies seem to support the claim that business simulation exercises trigger participants to do things differently, think longer term, see the big picture, and better understand the competitive landscape (Gwynne, 1995; Hequet, 1995; Lefebvre, 1997; McCune, 1998; McIlvaine, 1999; Reibstein & Chussil, 1997; E. Sherman, 1996; Stewart, 1997; Wilson & Condom, 1995). Unfortunately, empirical studies that actually attempt to measure the benefits of a simulation exercise are limited and have produced mixed results (Anderson & Lawton, 1992; Chapman & Sorge, 1999; Keys & Wolfe, 1990; Randel, Morris, Wetzel, & Whitehill, 1992; Wolfe, 1985). Whereas previous research has focused on measuring the perceived benefits, the research presented in this article empirically exposes the process that drives these benefits.

Justification for the study

This study provides insight into what happens to participants in a business simulation exercise. Based on experience, several researchers have documented the benefits

AUTHOR’S NOTE: I would like to thank the anonymous reviewers for their excellent comments and suggestions.
of these exercises (Chapman & Sorge, 1999; Lefebvre, 1997; Reibstein & Chusil, 1997). The most common reported benefits include practice in an environment without risk, increased creativity, decreased decision time, more focused competitive analysis, increased cross-functional understanding, and increased subject content knowledge. Although case studies and practitioner experience support these benefits, little empirical evidence is offered in the training literature on the change process that an individual participant experiences.

Furthermore, there is the question of external validity, or whether the benefits identified will be transferred to the actual business environment. Previous studies by Wolfe and Roberts (1986, 1993) supported the external validity of business simulation exercises, whereas a study by Norris and Snyder (1980) refuted these findings. A key element in resolving the contradictory results is to understand the process that brings about the measured benefits. Do the benefits result in a quantifiable change in the individual underlying decision process? Are the changes the same for each individual?

The first step in addressing these difficult questions was to measure whether a change has taken place in the individual exposed to the business simulation exercise. Has the business simulation exercise in some way changed the perspective of participants? As Evans and Wurster (2000) pointed out, the biggest risk today for businesses is not “legacy assets” but the “legacy mindsets” (p. 66). In a landmark book, Creativity, Csikszentmihalyi (1996) reported the results of interviewing more than 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 such as Einstein’s theory of relativity were the result of thinking differently (H. Sherman & Schultz, 1998). These thinkers not only changed their own perspectives, but they created new perspectives for others to follow. If the perspectives of people could be changed, all indications were that individuals would think differently, and creative new approaches would emerge. Thus, changing individual perspectives should be an objective of management training and business simulation exercises. In addition, measuring this change should be a central research focus as a principal indicator of a training exercise’s value.

Objectives and hypotheses

The experimental objective of this study was to test a number of simulation design hypotheses regarding the changes that would be induced in an individual’s mental models. This objective was based on the assumption that participants completing a business simulation exercise would characterize their decisions differently after the exercise and would identify new solution approaches to these same decisions. For example, it could be hypothesized that a decision problem that was characterized as being ordinally more complex and intuitive prior to the business simulation experience would be ordinally simpler after the training and the solution approach taken would be ordinally more textbook. Generically, the hypotheses were presented as fol-
Hypothesis Form 1: Decision target \(X_n\) is characterized as ordinally more/less \(C\) after participation in the business simulation exercise.

Hypothesis Form 2: Decision problem \(X_n\) is approached ordinally more/less \(A\) after participation in the business simulation exercise.

The specifics of these generic hypotheses that define the target decisions will be developed and detailed in the next several sections. The rejection or failure to support these hypotheses will be used to identify where significant changes have occurred and to assess the impact of a business simulation exercise. The experimental objective could be simply stated as finding the supporting evidence to accept targeted hypotheses based on simulation objectives of the forms 1 and 2.

Method

This research applied the semantic differential technique to quantify how business simulation exercises could change a decision maker’s mental models and perceived decision-making actions. The semantic differential technique was used to understand a business simulation decision process and expose its underlying meaning. By capturing the decision problems underlying dimensional meaning and measuring this meaning in a pretest-posttest experimental design, changes in a business simulation exercise participants’ mental models could be studied, and the benefits of the simulation could be empirically evaluated.

Setting

The participants in the study were a relatively homogeneous group of 21 senior vice presidents from a medium-sized service business. The participants all had a high level of understanding of their particular industry’s dynamics and were familiar with the decisions that drive business profitability. All participants performed the same job function for their company and had similar responsibilities. The participants were divided into four teams by their firm’s nonparticipating executive management, and then they made decisions in a simulated business environment.

The intervention was the custom business simulation called CYCLOAN (Lefebvre & Scherpereel, 1998) developed by PriSim Business War Games Incorporated. CYCLOAN provided a competitive rehearsal simulation environment for the teams of participants to run the branch office of a service company. The participant teams made decisions that were both commonly encountered in the management of a mortgage bank branch office and identified as key drivers of business success. Performance in any particular year was based on a model of typical industry dynamics and the competitive landscape created by other participant teams. Facilitators (nonparticipants) were
available to answer questions regarding the simulated environment and the decisions that needed to be made.

**Research design**

A business simulation exercise was used to examine the independent variables to be tested in this study, namely (a) the decision characteristic and (b) the decision approach concepts. Although the independent variable was changed by the intervention, the measurement instrument attempted to confirm the hypothesized changes in the dependent variables. In this design, the dependent variables were the eight measurements (semantic differential scales) associated with each decision concept.

This study uses a preexperimental one-group pretest-posttest design (Campbell & Stanley, 1963). Only a single homogeneous group was subjected to the pretest measurement, business simulation intervention, and posttest measurement sequence suggested by the design. Limiting the design to the preexperimental grouping became a necessity because the target decisions were customized to the group being tested and the population was too small to allow for a reasonable control group. Thus, as discussed in more depth in a later section, the results might not be generalized easily to other groups.

Care was taken to ensure that the design was applied in a manner that maximized its power to measure the hypothesized relationships. Given the preexperimental design, efforts were focused on establishing high levels of internal validity. The group was 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 only 21 participants receiving the intervention provided some assurance that all participants experience the intervention equivalently.

A retreat type format was chosen for the business simulation exercise. This format allowed the participants to focus their energy on running a simulated business and not be distracted by daily job demands. Thus, the measured intervention effects could be better isolated from external factors. This format was also ideal for controlling for many of the parameters associated with the passage of time. The pretest measurement was taken at the beginning of the first day, and the posttest measurement was taken at the end of the third day. In the interim time, the participants were staying in a remote location and instructions were explicitly given to the participants by their executive management that “running the simulated business should be considered their top priority.”

**Decision selection**

The target decision concepts to study were selected considering the designer’s and sponsor’s objectives, the participant’s familiarity with the decision set, and the participant’s ability to rate these decisions along a variety of semantic scales. Based on learning studies of Chinese characters by Hull (1920), it was assumed that as an individual...
obtains greater experience with a particular decision, an identifiable and consistent mental model would emerge. By selecting common decisions, or decisions with which the participants were familiar, participants were able to recall an unambiguous mental concept that they could then use for rating the semantic differential scales. This reduced the major concern that participants would be rating the textual concepts on the semantic scales based on different mental concepts and increased the likelihood that the data could reliably be aggregated to generate the true concept meaning.

Another consideration was that the decisions selected were those made during the business simulation exercise. Because the intervention being used in the study, the CYCLOAN business simulation exercise, was designed with a limited number of decisions that focused participant learning, the set of possible decision concepts was constrained. The decisions in the custom exercise were those identified by the game designer, sponsor, and consultants as critical to the running of the business. These critical decisions were the ones targeted for change by the business simulation designers and therefore were the ones selected for measurement (Goosen, 2001).

Figure 1 contains a shortlist of concepts selected as potential candidates for measurement. Interviews with the business simulation exercise designers and sponsors reduced this shortlist to the six concepts highlighted in boxes. These six concepts cover a broad range of business decisions, from personnel management to business strategy. The range was selected to cover a variety of decisions commonly encountered by the participants in their actual work environments.

The six target decision concepts selected from Figure 1 were 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 abbreviations are used in the remainder of this article.

Instrument design

Exploratory and pilot studies were conducted to develop the final instrument design. Two sets (characterization and approach) of eight bipolar semantic differential scales were developed as the proxy measurements for each of the decision targets indicated in Figure 1. The decision dimension questionnaire began with a page of detailed
instructions describing the correct marking of the instrument. As suggested by a pilot study, the questions were 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 were presented so the same decision target that appeared in the decision characteristic column was mirrored in the decision approach column. Finally, space limitations allowed only two decision targets to be included on each page. The final layout is illustrated in Figure 2. Decision targets and measurement scales were presented in a sequence to reduce respondent bias (Emory, 1985).

### Hypotheses details

A hypothesis was formulated for each target decision and semantic differential scale combination, for a total of 96 specific hypotheses for each of the two generic hypotheses. The construction of each hypothesis was based on the objectives of the business simulation exercise. For example, it was hypothesized that after completing the simulation exercise, the target decision establishing strategy would be characterized as more ambiguous and would be identified as requiring a more gut approach. The objective of the business simulation exercise was to identify applicable strategic techniques that would change the perspective on the target decision toward clear and the participant’s thinking toward textbook. Figure 3 presents 12 of the highly targeted changes identified by the simulation designers and sponsors.

The hypotheses of form 1 and form 2 attempted to confirm that a change in decision perception and a change in thinking had occurred in the simulation exercise participants. This confirmation was one indicator that the participant’s reaction in an actual decision situation would be different as a result of the exercise. As shown in prior research, a connection exists between perception and response (Csikszentmihalyi, 1996; Evans & Wurster, 2000; H. Sherman & Schultz, 1998). In the context of this study, the interest was in the relationship between perspective and decision (a response to a situation) and using a business simulation exercise to improve the decision making of the participants.
Prior research by Klein (1993) highlighted the relationship between perspective and decision. Klein conducted a study that analyzed data from different domains and more than 600 decision points to identify the sources of decision errors. The following three sources emerged from these data: lack of experience, lack of information, and explaining away. The third source, explaining away, was identified as the result of a perspective paradigm. The decision maker had a mental model, or perspective, of the decision that was difficult to change. The semantic differential technique has the capability to reveal a decision maker's mental model and monitor any changes in perspective that occur. Detection of a perspective paradigm shift is the focus of the two sets of hypotheses of the forms 1 and 2.

**Data preparation**

After verifying completeness, responses were coded and labeled as indicated in Figure 4. For example, a question was labeled H/S/C – to quickly communicate that the decision target was hiring, the semantic scale was the decision characteristic scale simple-complex, and the scale’s origin was the pretest questionnaire. In this study, there were no cases of missing data; thus, data editing focused 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 relied on the subjective inspection of the questionnaires to identify cases where a particular scale was marked at the same level for every decision target or a long sequence of scales were marked at the same level. No abnormalities were noted in the data set.

---

**FIGURE 3:** Hypotheses for Targeted Changes

<table>
<thead>
<tr>
<th>Independent Variable (Decision Target)</th>
<th>Dependent Variable</th>
<th>Direction of movement*</th>
<th>Reasoning</th>
</tr>
</thead>
</table>
| Characteristic
| Hiring |
| Training |
| Time Blocking |
| Strategy |
| Sourcing |
| Approach |
| Hiring |
| Training |
| Time Blocking |
| Strategy |
| Sourcing |
| Notes |

* 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 BSE.*

---

* Key

- `<` significant movement toward left pole
- `>` significant movement toward the right pole
- `<<` significant movement but no direction specified
- `>>` no change anticipated
- BSE Business Simulation Exercise
total of 21 matching pretest and posttest questionnaires were collected, representing a 100% response rate.

Limitations of methodology and research design

Given the field research nature of this study, several limitations centering on the study’s reliability and validity are acknowledged. Although reliability and validity of the semantic differential technique have been confirmed in several studies (Bagozzi, 1981; Heise, 1970; Nickols & Shaw, 1964), the constraints relevant to this study introduce additional concerns.

Reliability is typically associated with the concept of stability. It is concerned with estimating the degree to which the measurement being taken is free from random or unstable error (Emory, 1985). 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 retesting control was not implemented in the methodology because the time of posttest was critical to obtaining an accurate measurement. It was 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 was an assumption based on the questionnaire’s careful construction. To enhance stability, the decision targets that were chosen for measurement all had high levels of familiarity among the participants. Thus, if environmental parameters could be controlled, there was little reason to believe that the respondents would alter their responses if the group were retested at a later point in time.

Reliability is a necessary but not sufficient condition for achieving internal and external validity. Internal validity is the extent that the differences being measured by the instrument reflect the true differences among those being tested. All three major forms of internal validity—content, criterion related, and construct—were considered
in the study design. The first, content validity, was addressed by selecting six of the major decisions contemplated by the participants in the business simulation exercise. This selection was determined by the exercise developers and the participants’ executive management to cover a broad range of familiar business decisions. Second, the extensive scale development effort provided some assurance that the criteria were valid. The scale identifiers used for measuring the different decision targets were selected to mirror the generic descriptions of these decisions. Finally, construct validity was confirmed by checking consistency among multiple decision targets in preliminary testing. In this testing, measurements on similar decision targets correlated in predictable ways.

Probably the greatest limitations of this study were related to external validity. Two external validity issues with the sample tested were apparent; the sample consisted of a relatively small number of participants and the participating group was homogeneous. The small sample size was a constraint introduced by the nature of the field intervention. Business simulation exercises in an executive environment are typically delivered to groups of between 12 and 36 individuals. Thus, the measurement of effect from a particular business simulation intervention was constrained to a relatively small sample. Attempting to aggregate data from multiple small samples and draw conclusions would endanger the study’s reliability because the objectives established for the particular business simulation exercise were group-specific. This unfortunate trade-off restricted the analytical techniques to less powerful nonparametric statistics on a small sample and put at risk the external validity of the findings.

The homogeneity of the group also limits the external validity of the study. When business simulation exercises are conducted in a corporate environment, the participants typically have similar educational, social, and personality types. Thus, the ability to generalize the findings beyond the specific group being studied, or even to groups with similar characteristics, would be suspect. Again, the external validity was weakened to maintain reliability and to facilitate greater understanding of the phenomena in question. As Snow and Anderson (1991) discussed, the criteria for generalizability depends on whether the conclusions seem reasonable to a broad cross-section of readers. This criterion of face validity seems evident in the findings presented in the following.

**Data analysis and findings**

Support was sought for two concept change hypotheses. Although parametric analysis techniques are justified for use on semantic differential data sets (Emory, 1985; Kerlinger, 1973), the small sample limitations of this study require the use of the less powerful nonparametric statistics. Snider and Osgood (1969) indicated that additional support for the hypotheses can be obtained using a frequency heuristic of individual cases. The combined evidence of the two techniques significantly strengthens the supporting evidence for the two concept change hypotheses. Although 96 hypotheses were generated for each of the two generic hypotheses, for a total of 192 hypotheses,
The simulation was designed to induce the greatest changes along the targeted 12 concepts identified in Figure 3. With the limited sample size and weak power of nonparametric tests, only the targeted 12 concepts are presented in this article.

The first indication of change in meaning for the decision concepts comes by observing the gap data. If a gap median was significantly different from zero in absolute value, then a change had occurred. The question is at what value different from zero is the change meaningful? In other words, at what gap value is the statement confirmed that the concept’s meaning has changed? This question can be answered using the nonparametric Wilcoxon signed-rank test on the raw pretest and posttest data files.

The Wilcoxon hypothesis-testing procedure was developed to assess whether a sample’s median failed to meet a specified value. By examining the difference, or gap, data between the matched individual pretest and posttest scores, a nonparametric hypothesis test similar to the matched pairs t test was performed. The null hypothesis was that no difference in the medians existed between the pretest and posttest samples, or H0: median gap = hypothesized median gap = 0. The alternative hypothesis (HA) is that there was a significant gap between the pretest and posttest samples, with the difference hypothesized as either greater than, less than, or not equal to 0. These results were tabulated for the targeted 12 concepts in Table 1.

The Wilcoxon signed-rank test confirms 9 of the targeted 12 hypotheses at the 90% confidence level or above, indicating that the study participants’ perceptual characterization and selected approaches changed in the hypothesized direction. Specifically, the test confirmed five of the six desired decision characterization changes, with participants 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. Although the test does not confirm a change in the characterization of the hiring decision along the small-big dimension, the ordinal descriptor has changed in the hypothesized direction.

Also confirmed were four 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. Not confirmed were the hypotheses that the strategy decision would be more textbook and the sourcing decision would require a bigger picture approach. The two approach changes that could not be confirmed by the nonparametric test both had ordinal descriptor changes in the direction hypothesized.

The gap data can also be used to identify specific changes in an individual. A nonstatistical heuristic measure is simply the frequency or count of values exceeding certain changes. As Snider and Osgood (1969) stated, “The evidence shows that for individual subjects a shift of more than two scale units probably represents a significant change or difference in meaning.” They went on to add that “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, the small sample size, and the specificity of the concepts, it can be argued that the data collected in this study were group data or at least somewhere between the two extremes outlined by Snider and Osgood. Thus, this study conservatively considered
<table>
<thead>
<tr>
<th>Scale</th>
<th>Pretest Ordinal Descriptor</th>
<th>Posttest Ordinal Descriptor</th>
<th>Wilcoxon Statistic</th>
<th>Median Gap</th>
<th>p Value (=)</th>
<th>p Value (&lt;)</th>
<th>p Value (&gt;)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision characteristic concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H S/B</td>
<td>Quite+</td>
<td>Quite+</td>
<td>32.5</td>
<td>0.0</td>
<td>0.65</td>
<td>0.71</td>
<td>0.32&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>TR S/B</td>
<td>Quite+</td>
<td>Extremely+</td>
<td>54</td>
<td>-0.5</td>
<td>0.07&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.97</td>
<td>0.03&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>TB S/L</td>
<td>Slightly+</td>
<td>Quite+</td>
<td>84.5</td>
<td>-0.5</td>
<td>0.05&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.98</td>
<td>0.02&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>ST C/A</td>
<td>Slightly-</td>
<td>Quite-</td>
<td>26.5</td>
<td>0.5</td>
<td>0.11</td>
<td>0.06&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.95</td>
</tr>
<tr>
<td>SO C/C</td>
<td>Extremely+</td>
<td>Quite+</td>
<td>5</td>
<td>0.5</td>
<td>0.03&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.01&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.99</td>
</tr>
<tr>
<td>TA C/C</td>
<td>Extremely+</td>
<td>Quite+</td>
<td>13</td>
<td>0.5</td>
<td>0.15</td>
<td>0.08&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.94</td>
</tr>
<tr>
<td>Decision approach concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H D/B</td>
<td>Neutral</td>
<td>Quite+</td>
<td>78.5</td>
<td>-1.0</td>
<td>0.11</td>
<td>0.95</td>
<td>0.06&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>TR D/B</td>
<td>Neutral</td>
<td>Quite+</td>
<td>41</td>
<td>-1.0</td>
<td>0.03&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.99</td>
<td>0.02&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>TB D/B</td>
<td>Slightly-</td>
<td>Quite+</td>
<td>118</td>
<td>-1.5</td>
<td>0.03&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.98</td>
<td>0.03&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>ST T/G</td>
<td>Slightly-</td>
<td>Quite-</td>
<td>44</td>
<td>0.5</td>
<td>0.22</td>
<td>0.11&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.90</td>
</tr>
<tr>
<td>SO D/B</td>
<td>Slightly+</td>
<td>Quite+</td>
<td>84.5</td>
<td>-0.5</td>
<td>0.41</td>
<td>0.81</td>
<td>0.20&lt;sup&gt;d&lt;/sup&gt;</td>
</tr>
<tr>
<td>TA D/B</td>
<td>Slightly+</td>
<td>Quite+</td>
<td>57.5</td>
<td>-0.5</td>
<td>0.16</td>
<td>0.93</td>
<td>0.08&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

NOTE: Abbreviations used in table are detailed in Figure 4. For example, H S/B represents the small/big characteristic scale for the decision target hiring a loan officer. 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. A priori hypothesis.

b. 90% confidence level.

c. 95% confidence level.
all changes that equal or exceed two scale units as significant when analyzing individ-
ual cases. Table 2 contains these data for the targeted 12 semantic differential scales.

A conservative look at the changes in individual cases, as shown in Table 2, indicates significant changes occurred for specific decision concepts in 9.5% to 61.9% of the participants. Although not all of these changes were in the directions hypothesized, a majority confirmed the hypotheses, where 9.5% to 47% of the participants exhibited significant changes in the hypothesized direction. From a simulation designer and manager perspective, the significant results presented in Table 2 are the most useful for adjusting the business simulation exercise to have greatest individual impact. Although these results provide additional confirming evidence for the hypotheses, they introduce some concern for both the simulation designer and the simulation sponsor.

Why did several individuals indicate a significant perceptual change counter to the direction hypothesized? Is this a serious problem with the simulation exercise or with the measurement methodology? The CYCLOAN business simulation exercise was designed to change participants’ perspectives, thus any significant change regardless of direction would indicate a change in perspective and meet the simulation objectives. Because the changes being identified were not learning objectives but rather indicators of perspective changes, a movement counter to the hypothesized direction should not be considered a failure of the simulation or measurement methodology. Because every individual’s experience and interaction with the simulation environment was unique, it might not be possible to prescribe a particular perceptual change uniformly across all individuals. However, as a group, the perceptual change seems relatively predictable, and the indicators developed in this research could be used to guide modifications to the simulation to improve the consistency of response.

Conclusions and future research

Business simulation exercises are designed to achieve specific objectives. Thus, it was the simulation exercise designer’s goal to capture the sponsor’s requirements and, based on these requirements, develop an exercise that induced targeted change in the participants. This study identifies six key decision problems designed into a business simulation exercise and formulated hypotheses regarding the expected change induced in the participants. The Wilcoxon signed-rank test confirmed 9 of 12 key hypotheses. Anecdotal and heuristic evidence offered additional confirming support. Given the evidence, the measurement methodology developed provides quantifiable justification for the use of business simulation exercises to induce targeted change in a decision maker’s mental models. In addition, business simulation designers and spon- sors can apply the methodology presented to confirm that the desired simulation objectives have been achieved. They can then intelligently adjust the exercise’s parameters to perhaps enhance the changes already observed. Future research will provide support for the usefulness of this feedback in improving business simulation effectiveness.
### TABLE 2: Individual Cases—Summary of Change

<table>
<thead>
<tr>
<th>Scale</th>
<th>Hypothesized Change</th>
<th>Percentage With Significant Change (≥ 2 units)</th>
<th>Count With Significant Change</th>
<th>Percentage With Significant Change in Hypothesized Direction (≥ 2 Units)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision characteristic concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H S/B</td>
<td>&gt;</td>
<td>2</td>
<td>9.5</td>
<td>2</td>
</tr>
<tr>
<td>TR S/B</td>
<td>&gt;</td>
<td>4</td>
<td>19.0</td>
<td>4</td>
</tr>
<tr>
<td>TB S/L</td>
<td>&gt;</td>
<td>7</td>
<td>33.3</td>
<td>6</td>
</tr>
<tr>
<td>ST C/A</td>
<td>&lt;</td>
<td>9</td>
<td>42.9</td>
<td>7</td>
</tr>
<tr>
<td>SO C/C</td>
<td>&lt;</td>
<td>6</td>
<td>28.6</td>
<td>6</td>
</tr>
<tr>
<td>TA C/C</td>
<td>&lt;</td>
<td>6</td>
<td>28.6</td>
<td>5</td>
</tr>
<tr>
<td>Decision approach concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H D/B</td>
<td>&gt;</td>
<td>10</td>
<td>47.6</td>
<td>7</td>
</tr>
<tr>
<td>TR D/B</td>
<td>&gt;</td>
<td>8</td>
<td>38.1</td>
<td>7</td>
</tr>
<tr>
<td>TB D/B</td>
<td>&gt;</td>
<td>13</td>
<td>61.9</td>
<td>10</td>
</tr>
<tr>
<td>ST T/G</td>
<td>&lt;</td>
<td>7</td>
<td>33.3</td>
<td>5</td>
</tr>
<tr>
<td>SO D/B</td>
<td>&gt;</td>
<td>8</td>
<td>38.1</td>
<td>5</td>
</tr>
<tr>
<td>TA D/B</td>
<td>&gt;</td>
<td>9</td>
<td>42.9</td>
<td>6</td>
</tr>
</tbody>
</table>

NOTE: Abbreviations used in table are detailed in Figure 4. For example, H S/B represents the small/big characteristic scale for the decision target hiring a loan officer.
The techniques developed in this research were applied to a real-world problem of evaluating the benefits of a business simulation exercise. The decision measurement technique was able to clearly demonstrate that a business simulation 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 a business simulation exercise was able to meet specific objectives provided insight into the process that brings about the observed benefits of simulation reported in previous studies. The technique’s success provides quantitatively measured justification for a business leader sponsoring a business simulation exercise that is intended to change people’s perspectives.

The measurement methodologies developed specifically to quantify the value of a business simulation exercise can be extended to other training and development activities. The methodologies developed in this research are especially applicable to experiential learning, 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 experiential activity. This contribution is significant for both decision makers, who are attempting to justify expenditures on these programs, and training firms, which are developing and promoting the use of experiential exercises.

The empirical evidence presented in this research had a number of limitations. The sample size was small and was selected from a single homogeneous group of decision makers. These limitations leave open a number of what if questions. What if more decision makers were studied, would the results be the same? What if a different set of decision makers were studied? What if the group studied were 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 would also allow more sophisticated techniques to be used to statistically model the data structure.

As currently developed, the measurement technique offers a tool for identifying, measuring, communicating, and targeting change in an organization’s people. By exposing the underlying process and individual’s mental models, the changes that business simulation exercises produce can be better understood. In addition, feedback can be communicated to the simulation designer so that design objectives can be validated and, if necessary, the design can be modified.

References


Christopher M. Scherpereel is an assistant professor of management at Northern Arizona University. He holds degrees from Notre Dame, Georgia Institute of Technology, New York University, and received his Ph.D. in industrial engineering and management science from Northwestern University in 2001. His research interests include decision making and analysis, real options, entrepreneurship, simulation, and strategy.

ADDRESS: CMS: Assistant Professor, Management, Northern Arizona University, College of Business Administration, PO Box 15066, Flagstaff, AZ 86011-5066, USA; telephone: +1 (928) 523-7831; fax: +1 (928) 523-7331; e-mail: chris.scherpereel@nau.edu.