Using Causal Path Analysis to Test & Validate Strategic Hypotheses

Dr. Palmer Morrel-Samuels
University of Michigan
Employee Motivation & Performance Assessment, inc.

Paper presented at Kaplan & Norton’s conference in Cambridge, Massachusetts Making Strategy a Continuous Process through the Balanced Scorecard, 12/16/99
Agenda

I) The Role of Survey Validity in Strategic Planning

II) A General Method for Synthesizing Soft Voices & Hard Numbers

III) Causal Path Analysis -- Strengths & Weaknesses of 3 Methods

IV) Case Profiles: Examples of the Rigorous Quantitative Approach
Section I

The Role of Survey Validity in Strategic Planning
Proof of Poor Validity - Standard Deviations & Means are Inconstant
Validity - Standard Deviations and Means are Flat Throughout the Survey
The Diamond Diagnostic gives significantly higher Leadership scores to executives who subsequently come within 10% of their financial target.
Validity - Data are Distributed Normally in the Dataset
Validity - Perceived Quality Predicts Prevention of Hospital Patient Falls

Perceived Quality of Care (Q23-Q)

FALL PREVENTION

![Graph showing the relationship between perceived quality of care and fall prevention.](image-url)
Validity - Leadership Predicts Prevention of Hospital Employee Injuries

Mean(S1: AVE LEADERSHIP & MGT)

Mean (INJURY PREVENTION)
Validity - Quality Predicts Prevention of Post-treatment Readmission

![Graph showing the relationship between Perceived Quality of Care and Readmission Prevention. The graph displays a positive correlation, with higher perceived quality correlating with higher readmission prevention.]
Validity - Communication Predicts Hospital Patient Satisfaction
Validity - Estimated Employee Motivation Tracks Actual Motivation
Employees with a high Employee Motivation Index write survey comments that are more positive.
Validity - Summary Questions Confirm Behavior Questions

The high correlation proves that the survey asked the right questions in each of the topic areas.
Validity - Truck Defect Rate Predicts Customer Satisfaction
Validity - Good Management Predicts Low Factory Absenteeism
Validity - Less Awkward Jobs Have Lower Accident Costs
Validity - High Motivation Index Scores Predict **Low Defect Rate**
Poor survey data will give you poor predictions every time.
Section II

A General Method for Synthesizing Soft Voices & Hard Numbers
A Straightforward Process

1) Solicit voice of the customer and/or the employee with a focused survey
   a) Use qualitative & quantitative methods
   b) Preserve anonymity & confidentiality
   c) Include an indexing variable (e.g., SU, or job, or region)

2) Compile quantitative data into a quantitative database
   a) One row for each respondent, 1 column for each question
   b) Include the chosen indexing variable
   c) Import hard performance metrics on profit, turnover, defect rate, etc.

3) Code qualitative data and enter into a qualitative database
   a) Use double blind procedures wherever possible
   b) Limit number of categories in each constructed variable
   c) One row for each interviewee
   d) Include the chosen indexing variable
   e) If each level of the indexing variable has more than 1 row, then make
      a transitional database that concatenates or counts or gives a ratio, so
      that each level has no more than 1 row

4) Merge qualitative & quantitative databases for statistical analysis
Section III

Causal Path Analysis;
Strengths & Weaknesses of 3 Methods
### Method 1: Comprehensive Inter-Correlations

“Managers can help validate hypothesized cause-and-effect relationships by measuring the correlation between two or more measures.”

Kaplan & Norton, The Balanced Scorecard, 1996 pg. 254

![Table with correlations]

**One simple zero-order correlation**

<table>
<thead>
<tr>
<th></th>
<th>Work Conditions</th>
<th>Team Work</th>
<th>Communication</th>
<th>Leadership</th>
<th>Blame</th>
<th>Team</th>
<th>Skill</th>
<th>Tool</th>
<th>Focused</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
<td>NEG</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>Gross Profit</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>Net Income</td>
<td>POS</td>
<td>NEG</td>
<td>POS</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>G/P per Total Expense Dollar</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>G/P per EE</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>Net Income per EE</td>
<td>POS</td>
<td>NEG</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>G/P per Comp Dollar</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>Net Income per Comp Dollar</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>Staff Retention %</td>
<td>NEG</td>
<td>POS</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
<tr>
<td>Staff Growth %</td>
<td>NEG</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
<td>POS</td>
</tr>
</tbody>
</table>
Benefits & Disadvantages - Inter-correlations

Benefits:
- Ease of computation
- High clarity during rollout

Disadvantages:
- Possible errors due to spurious correlation
- Possible omissions due to masked effects
Simple Plots May Mask Actual Associations

No correlation appears between Sales per Employee & Receptivity to Change...
Simple Plots May Show Spurious Correlations

and Sales per Employee appears to drop as Number of Employees rises. However...
HOWEVER...

Receptivity to Change **DOES** emerge as a consistent driver for Sales per Employee, but only when we control for all available demographic covariates (viz.: Region, Number of Employees, Company Age & Company Type.) In the full partialled analysis only Company Type & Receptivity to Change are significant drivers.

Leverage Plots show Association Controlling for Confounding Variables

All prescription drugs in the US are tested with this type of statistical analysis.
“Exploring the complex dynamics will likely require simulation and cost modeling.” K&N HBR ’92, p. 79. See also Fortune, 10/13/97: Bringing Sears into the New World p. 183-184: “Now we know that if a store increases its employee satisfaction score by five measuring units this quarter, the following quarter its customer satisfaction scores will go up by two units. And if a store increases its customer satisfaction by two units, its revenue growth the following quarter will beat our stores’ national average by 0.5%” p. 184. Customer Satisfaction Model for Cathay Pacific Airline, Business Class; adapted from Customer Orientation & Market Action, M. Johnson, p. 122
Benefits & Disadvantages - Causal Modeling with PLS

Benefits:
Models can accommodate complexity even with a few Latent Variables (LVs) (H, ‘88, p. 221)
Can run well even with small datasets of non-normal distributions (FT&Z ‘82, p. 405)
Can be used well even in datasets where multicollinearity is high (P&R, ‘95, p. 184 ff.)
Does not require ratio scales with an absolute zero and even intervals (P&R, ‘95, p. 191)

Disadvantages:
Results are model-specific: Excluding a link or Manifest Variable (MV) can ruin all predictions*
Comparative testing of alternate models is not straightforward (G ‘94, p. 124)
Magnitude of linkage is usually underestimated (WRW&D, ‘84, p. 740)
Valence of linkages is unstable (WRW&D, ‘84, p. 741)
MVs must be standardized, but standardization makes coefficients unstable (C&C, ‘83, p. 366)
Direction of MV-LV arrow is theory-determined but can effect results (F&C, ‘94, p. 75)
Predictive power may be very low even in good models (J, ‘98, e.g., 123)
Results very unstable where many non-essential factors have small impacts (H&A, ‘94 p. 589)
Conventional significance tests are approximate and/or problematic (H ‘88, p. 222 ff.)
LVs are necessarily independent, so their scores and interrelations may be unrepresentative

* “In cases in which the causal relations are uncertain, the method [path analysis] can be used to find the logical consequences of any particular hypothesis in regard to them.” (W, 1921, p. 557)
# Method 3: Hierarchical Modeling (used at Disney)

<table>
<thead>
<tr>
<th>Work Conditions</th>
<th>Teamwork</th>
<th>Communication</th>
<th>Leadership</th>
<th>Management</th>
<th>Team Focus</th>
<th>Skill Focus</th>
<th>Tool Focus</th>
<th>Salary Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sales</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gross Profit</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net Income</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G/P per Total Expense Dollar</td>
<td>Promotes</td>
<td>Promotes</td>
<td>Promotes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G/P per EE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net Income per EE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G/P per Comp Dollar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Net Income per Comp Dollar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff Retention %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Staff Growth %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **r zero-order (disaggregated n=438)**
- **r 2 (correlational model 2R)**
- **r 3 (correlational model 3R)**
- **Beta 1 (main effects model)**
- **Beta 2 (model of profit alone)**
- **Beta 3 (model number 3B)**
- **Beta 4 (model number 4B)**
- **MANOVA 1 (main effects model)**
- **MANOVA 2 (model number 2M)**
- **MANOVA 3 (model number 3M)**
Benefits & Disadvantages - Hierarchical Modeling

Benefits:
Results hold *regardless* of statistical model used
Good clarity during rollout
Predictions are conservative (low Type I Error)
Results are highly-defensible

Disadvantages:
Computational rules are required to tally results
Section IV

Case Profiles: Examples of the Rigorous Quantitative Approach
An analysis of linkages in ‘93 showed that Leadership drove Quality in this division of GTE. Programs to improve Leadership were implemented over the next two years, and a customized employee survey documented their effectiveness: Leadership rose each year. As predicted, when Leadership rose so did Quality: Defect Rate dropped by 22% in 1994 and by an additional 24% in 1995.
In a recent engagement, EMPA designed a customized annual survey for the 20K employees of XYZ. The linkage analysis predicted that increasing what the employee survey identified as Career Focused Workstyle would boost financial productivity regardless of the metric used. After the survey, XYZ rolled out a set of programs to help boost Career Focused Workstyle at a cost of about $60K. EMPA’s next annual survey tracked the change in the corporate culture as workstyle increased by 5%.

Multivariate statistical analysis also proved that this 5% increase in Career Focused Workstyle generated $3.3M in new revenue for the company.
Quantitative Analysis of Linkages Improves Quality at INS

Workforce: 20,000 employees in 350 Ports-of-Entry across the nation at our borders

Problem & Scope: No objective QC for 500,000,000 INS Inspections per year

Intervention: Quality Checks of inspected & approved travelers before entry to the US

Method: A Multiphase-multistage Proportional Stratified Random Sample

Construct Validity: Quality Checks show no bias for Gender, Age, or Citizenship

Predictive Validity: Good Communication predicts low Defect Rate in Inspections

Content Validity: High Commitment yields high Thoroughness in the Quality Check

Business Utility: Catch Rates vary at the POEs, so strategic interventions are possible
# References

<table>
<thead>
<tr>
<th>Author(s)</th>
<th>Title</th>
<th>Publication Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heskett et al.</td>
<td>Putting the Service-Profit Chain to Work</td>
<td>HBR, March-April 1994</td>
</tr>
<tr>
<td>McBurney, O’Neill &amp; Mills</td>
<td>Ch. 7: Evidence, Argumentation and Debate</td>
<td>NYC: Macmillan</td>
</tr>
<tr>
<td>Kaplan &amp; Norton</td>
<td>The Balanced Scorecard</td>
<td>HBR, Jan-Feb</td>
</tr>
<tr>
<td>Cohen &amp; Cohen</td>
<td>Applied Multiple Regression/Correlation...</td>
<td>Hillsdale: Erlbaum</td>
</tr>
<tr>
<td>Klein et al.</td>
<td>Race &amp; Imprisonment Decisions in California</td>
<td>Science, 247 (4944)</td>
</tr>
<tr>
<td>SAS</td>
<td>Comparison of Means, SAS/STAT User’s Guide</td>
<td>v 6, 2, Cary: SAS</td>
</tr>
<tr>
<td>Morwitz, Johnson &amp; Schmittlein</td>
<td>Does measuring intent change behavior?</td>
<td>Journal of Consumer Research, 20 (1) 46-61</td>
</tr>
<tr>
<td>Abelson</td>
<td>Statistics as principled argument</td>
<td>Hillsdale: Erlbaum</td>
</tr>
</tbody>
</table>


Helland & Almøy (1994) Comparison of prediction methods when only a few components are relevant, Journal of the American Statistical Assoc., 89, (426) 583-591.

Questions/Comments: Palmer Morrel-Samuels
EMPA inc., 2155 Jackson Ave. Ann Arbor, MI 48103
Tel: (734) 327-8800 palmer@umich.edu
www.SurveysForBusiness.com