12-2002

The Effects of Group Size on Incentive Effectiveness: A Meta-Analysis

Angelica C. Grindle
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/dissertations

Part of the Applied Behavior Analysis Commons, and the Experimental Analysis of Behavior Commons

Recommended Citation
https://scholarworks.wmich.edu/dissertations/1229

This Dissertation-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Dissertations by an authorized administrator of ScholarWorks at WMU. For more information, please contact maira.bundza@wmich.edu.
A meta-analysis was conducted to examine the effects of group size on incentive effectiveness using data obtained from incentive systems implemented in 13 companies. Other predictor variables included the organizational level targeted for improvement, long/short-term profitability type, company number, and length of exposure to the incentive system. The main dependent variable was the monthly score for each measure of performance. Three types of meaningful comparisons were defined for these data: (1) Between-Group – Within-Company comparisons in which a performance measure was in place in two or more units of an organization; (2) Between-Group – Across-Company comparisons in which a performance measure existed in two or more companies; and (3) Group Expansion comparisons in which a performance measure was in place in one unit of the organization and the group size increased over time. A variety of statistical analyses were conducted to examine the effects of the main independent variable (group size) and the other predictor variables on the performance scores. The results of this study are consistent with previous research as well as theory regarding the effects of group size on incentive effectiveness. That is, individual and small group incentives appear to be equally effective but this effectiveness decreases as the group size increases. Suggestions for future research are discussed.
ACKNOWLEDGMENTS

My sincere appreciation goes to Dr. Bradley Huitema for patiently guiding me through the completion of this dissertation. It would never have been achieved without him. Special thanks goes to Dr. Alyce Dickinson for the guidance and support she has provided to me throughout the years. I would also like to thank my other committee members Dr. John Austin and Dr. Joseph McKean. Lastly, my gratitude goes to Dr. William Abernathy of Abernathy and Associates for generously allowing me to delve into his databases.

I wish to thank my family for their unconditional love, support, and encouragement. Special thanks go to my parents, Don and Lucy Grindle, for always believing in me and to my sister, Jennifer Grindle, for providing a shining example of what it means to be a “Grindle Woman.” I am deeply grateful to my husband, Bill Boettcher, whose unfailing counsel and encouragement provided the foundation for the realization of this goal. Finally, my deepest gratitude goes to my daughter, Abigail, whose entry into the world has taught me about what is truly important.

Angelica C. Grindle
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vii</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Behavior Engineering Model</td>
<td>1</td>
</tr>
<tr>
<td>Monetary Incentives</td>
<td>4</td>
</tr>
<tr>
<td>Group Incentives, Individual Incentives, and Hourly Pay</td>
<td>5</td>
</tr>
<tr>
<td>Group Incentives vs. Hourly Pay</td>
<td>5</td>
</tr>
<tr>
<td>Individual Incentives vs. Hourly Pay</td>
<td>8</td>
</tr>
<tr>
<td>Group Incentives vs. Individual Incentives</td>
<td>8</td>
</tr>
<tr>
<td>Group Size</td>
<td>10</td>
</tr>
<tr>
<td>METHOD</td>
<td>13</td>
</tr>
<tr>
<td>Incentive System</td>
<td>13</td>
</tr>
<tr>
<td>Variables</td>
<td>15</td>
</tr>
<tr>
<td>Dependent Variables</td>
<td>15</td>
</tr>
<tr>
<td>Independent Variables</td>
<td>15</td>
</tr>
<tr>
<td>Group Size</td>
<td>15</td>
</tr>
<tr>
<td>Measure Level</td>
<td>16</td>
</tr>
<tr>
<td>Long/Short-Term Profitability</td>
<td>16</td>
</tr>
<tr>
<td>Length of Exposure</td>
<td>17</td>
</tr>
</tbody>
</table>

iii
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>17</td>
</tr>
<tr>
<td>Types of Comparisons</td>
<td>18</td>
</tr>
<tr>
<td>Analyses</td>
<td>20</td>
</tr>
<tr>
<td>Between-Group – Within-Company Comparisons of the Time/Performance Relationship</td>
<td>20</td>
</tr>
<tr>
<td>Time/Performance Relationship within Group Size Categories</td>
<td>22</td>
</tr>
<tr>
<td>Association with Predictors</td>
<td>22</td>
</tr>
<tr>
<td>Between-Group – Across-Company Comparisons of the Time/Performance Relationship</td>
<td>23</td>
</tr>
<tr>
<td>Group-Expansion Comparisons</td>
<td>23</td>
</tr>
<tr>
<td>Two-Parameter Model</td>
<td>23</td>
</tr>
<tr>
<td>Four-Parameter Model</td>
<td>26</td>
</tr>
<tr>
<td>Model Selection</td>
<td>27</td>
</tr>
<tr>
<td>Tests for Effects of Increase in Group Size</td>
<td>28</td>
</tr>
<tr>
<td>RESULTS</td>
<td>29</td>
</tr>
<tr>
<td>Between-Group – Within-Company Comparisons</td>
<td>29</td>
</tr>
<tr>
<td>Between-Group – Within-Company Differences in Time/Performance Relationship</td>
<td>31</td>
</tr>
<tr>
<td>Time/Performance Relationship within Group Size Categories</td>
<td>32</td>
</tr>
<tr>
<td>Association with Predictors</td>
<td>33</td>
</tr>
<tr>
<td>Between-Group – Across-Company Comparisons</td>
<td>35</td>
</tr>
<tr>
<td>Between-Group – Across-Companies Differences in Time/Performance Relationship</td>
<td>37</td>
</tr>
</tbody>
</table>
Table of Contents—continued

Time/Performance Relationship Within Group Size
Categories................................................................. 38

Group-Expansion Comparisons................................. 39

DISCUSSION ........................................................................................................ 43

APPENDICES

A. Between-Group – Within-Company Comparison Graphs .......... 48
B. Between-Group – Across-Company Comparison Graphs .......... 94
C. Group Expansion Comparison Graphs – Change Type One .......... 99

BIBLIOGRAPHY ................................................................. 124
# LIST OF TABLES

1. Six Components of Worthwhile Performance Example (Gilbert, 1978) ................................................. 2
3. Performance Scorecard Example ................................................................................................. 14
4. Cross-Classification of Company Variables and Company Code ................................................. 18
5. Number of Data Sets Associated with each Level of the Four Predictor Variables (Between-Group – Within-Company Comparisons) ......................................................... 29
6. Source Table for the ANOVA Conducted on the Average Fisher’s Z for the Individual, Small, and Large Groups (Between-Group – Within-Company) ........................................................................... 32
7. Multiple correlation coefficients (and Obtained p Values) for All Predictors in Each Group Size (Between-Group – Within-Company) ............................................................... 34
8. Obtained Part Correlation Coefficients (and p Values) from Part Correlation Analyses (Between-Group – Within-Company) ............................................................... 35
9. Number of Data Sets Associated with each Level of the Four Predictor Variables (Between-Group – Across-Company Comparisons) ................................................................. 36
10. Source Table for the ANOVA Conducted on the Average Fisher’s Z for the Individual, Small, and Large Groups (Between-Group – Across-Company Comparisons) ........................................................................... 38
11. Source Table for the ANOVA Conducted on the Standardized Effects for the One, Small, and Large Change Type Groups (Group-Expansion Comparisons) ........................................................................... 40
12. Results Table for the Kruskal-Wallis Conducted on the Standardized Effects for the One, Small, and Large Change Type Groups (Group-Expansion Comparisons) ........................................................................... 41
LIST OF FIGURES

1. The Three-Term Contingency ................................................................. 1
2. Between-Group – Within-Company Comparison Example .................... 19
3. Between-Group – Across-Company Comparison Example ..................... 19
4. Group-Expansion Comparison Example ................................................. 20
7. Average Weighted Fisher’s Z for each Group Size (Between-Group – Within-Company) .............................................................. 31
8. Average Fisher’s Z for each Group Size (Between-Group – Across-Company Comparisons) ........................................................... 37
9. Dotplot of the Standardized Effects ....................................................... 39
10. Mean Standardized Effect by Change Ratio (Group-Expansion Comparisons) ................................................................. 40
11. Median Standardized Effect by Change Ratio (Group-Expansion Comparisons) ................................................................. 42
12. Dotplot of Group-Expansion Analyses Lag-1 Autocorrelation Coefficients ................................................................................. 42
13. Dotplot of Group-Expansion Analyses $Z_{hm}$ ........................................ 42
INTRODUCTION

Behavior Engineering Model

To increase the frequency of a behavior, the behavior (R) must be followed in time with the presentation of a reinforcer (S_R). Stimulus control is obtained when the presentation of this reinforcer only occurs in the presence of a particular stimulus (discriminative stimulus - \( S^D \)) and not in its absence. These three components of behavior can be diagrammed as shown in Figure 1.

\[
\begin{array}{c}
S^D \\
\text{Discriminative Stimulus}
\end{array} \rightarrow \begin{array}{c}
R \\
\text{Behavior}
\end{array} \rightarrow \begin{array}{c}
S_R \\
\text{Reinforcer}
\end{array}
\]

Figure 1. The Three-Term Contingency.

Human performance in the workplace is typically a combination of multiple behaviors. For example, the task of tightening a screw includes the behaviors of grasping the screw and screwdriver and turning the wrist. Regardless of the specific behavior, performance is a product of a person’s behavioral repertoire and the supporting environment (Gilbert, 1978). Both of these components are necessary for worthwhile performance. For example, a typist may type at a fast rate (behavioral repertoire). However, if his typewriter is not working properly (supporting environment), he will not create a quality product. Similarly, if a carpenter has access to the top-of-the-line tools of her trade (supporting environment) but does not know how to use them (behavioral repertoire) she will not be able to create quality cabinets.

Both the behavioral repertoire and the environmental supports of each of the components of behavior (i.e., discriminative stimuli, behaviors, and reinforcers) can be examined when analyzing human performance. For example, suppose a factory
worker is responsible for assembling plastic components that come toward her work area. Imagine that a flashing light indicates when the component falls into a holding mechanism and that the factory worker then places a second component on top of the first and pushes a button to meld the pieces together. As Table 1 shows, to perform her job successfully the following must occur:

- **Discriminative Stimulus (Information).** The light must be bright enough (Data) and she must be able to discriminate it from all the other flashing lights in the factory (Knowledge).
- **Behavior (Instrumentation).** The button must depress (Instruments) and she must have the physical dexterity to push it down (Response Capacity).
- **Reinforcer (Motivation).** The worker must be rewarded for her performance (Incentives) and have adequate motivation to do the job (Motives).

Table 1

<table>
<thead>
<tr>
<th>Components of Behavior</th>
<th>Repertory of Behavior</th>
<th>Environmental Supports</th>
</tr>
</thead>
<tbody>
<tr>
<td>S^D</td>
<td>R</td>
<td>S_r</td>
</tr>
<tr>
<td>Information</td>
<td>Instrumentation</td>
<td>Motivation</td>
</tr>
<tr>
<td>Data (flashing light)</td>
<td>Instruments (button)</td>
<td>Incentives (piece-rate)</td>
</tr>
<tr>
<td>Knowledge (perceives light)</td>
<td>Response Capacity (pushes button)</td>
<td>Motives (“motivated”)</td>
</tr>
</tbody>
</table>
Gilbert (1978) represented these performance factors in a six-celled matrix called the Behavior Engineering Model (BEM) as detailed in Table 2. As a diagnostic tool, the BEM allows the user to determine where to focus his performance improvement efforts. That is, while one can impact performance by manipulating any of the cells in the BEM, the goal of a performance engineer is to determine where to get the greatest impact on performance with the lowest cost. To achieve this goal, Gilbert (1978) proposed that performance problems should be diagnosed by first examining the environmental supports and then the evaluating the person’s repertory of behavior. This sequence is recommended because it is typically easier to change the former than the latter. Similarly, improving information is simpler to achieve than improving instrumentation, which in turn is easier to improve than motivation. Thus, performance issues should be analyzed by using the sequence suggested in the BEM: (1) Data, (2) Instruments, (3) Incentives, (4) Knowledge, (5) Capacity, and (6) Motives. The focus of the current study is the third cell (“Incentives”).
Table 2

<table>
<thead>
<tr>
<th>Environmental Supports</th>
<th>Information</th>
<th>Instrumentation</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Data</strong></td>
<td><strong>Instruments</strong></td>
<td><strong>Incentives</strong></td>
</tr>
<tr>
<td></td>
<td>• Relevant &amp; frequent feedback about the adequacy of performance</td>
<td>• Tools &amp; materials of work designed to scientifically match human factors</td>
<td>• Adequate financial incentives made contingent upon performance</td>
</tr>
<tr>
<td></td>
<td>• Descriptions of what is expected of performance</td>
<td></td>
<td>• Non-monetary incentives made available</td>
</tr>
<tr>
<td></td>
<td>• Clear &amp; relevant guides to adequate performance</td>
<td></td>
<td>• Career-development opportunities</td>
</tr>
<tr>
<td>Person’s Repertory of Behavior</td>
<td><strong>Knowledge</strong></td>
<td><strong>Capacity</strong></td>
<td><strong>Motives</strong></td>
</tr>
<tr>
<td></td>
<td>• Scientifically designed training that matches the requirements of exemplary performance</td>
<td>• Flexible scheduling of performance to match peak capacity</td>
<td>• Assessment of people’s motives to work</td>
</tr>
<tr>
<td></td>
<td>• Placement</td>
<td>• Prosthesis</td>
<td>• Recruitment of people to match the realities of the situation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Physical shaping</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Adaptation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Selection</td>
<td></td>
</tr>
</tbody>
</table>

Monetary Incentives

As indicated in Table 2, incentives are monetary or non-monetary (e.g., recognition, gifts, career advancement) rewards delivered contingent upon
performance. The remainder of this section reviews the literature on group versus individual monetary incentives.

Group Incentives, Individual Incentives, and Hourly Pay

Wages are disbursed in varied ways. Some of these make pay contingent upon performance and others do not. If an employee receives hourly pay, the amount received is not contingent upon any performance level above that needed to maintain employment. With individual and group incentives, however, the amount of pay received is contingent upon the performance level of the individual or the group to which the individual belongs. Because incentives provide a link between performance and pay it is often assumed that incentive pay systems will result in higher performance levels than do hourly pay systems. Correspondingly, it has been hypothesized that group incentives provide a weaker link than do individual incentives, because they do not provide a direct link between performance and amount of pay received. Thus, the argument is that individual incentives will result in higher performance levels than will group incentives. Studies that examine the relative effectiveness of hourly pay, individual incentives, and group incentives are discussed in the following sections.

Group Incentives vs. Hourly Pay

Five studies have compared both group and individual incentive systems with hourly pay systems (Allison, Silverstein & Galante, 1992; Farr, 1976; Honeywell-Johnson, 1997; Roberts & Leary, 1990; Smoot, 1997). All but the Allison et al. (1992) study were conducted in laboratory settings.
Farr (1976) compared the effects of hourly pay, individual incentives, and group incentives (both differentially-distributed and equally-distributed) utilizing a between-group experimental design. Members of 48 three-person groups sorted computer data cards with various combinations of holes punched in them. During an equally-distributed condition, group members earned equal amounts of incentives. During a differentially-distributed condition, the highest performer in the group received one-half of the available incentives, the second highest performer received one-third of the available incentives, and the lowest performer received the remaining one-sixth of the available incentives. Both types of group incentives resulted in higher performance levels than did hourly pay.

Smoot (1997) compared the effects of flat rate, individual incentives, and equally divided group incentives by utilizing a multiple baseline design. Members of six three-person groups assembled pop bead widgets. In addition, the researchers examined three types of pay scales: linear, positively accelerating, and negatively accelerating. In the linear pay scale condition, participants earned a set amount per widget. In the positively and negatively accelerating conditions, the amount of incentive increased or decreased, respectively, as production increased. Irrespective of the type of pay scale, the group incentive condition resulted in higher levels of performance than did the flat rate condition.

Roberts and Leary (1990) used a single-subject withdrawal design to compare performance under hourly pay, individual incentive, and group incentive conditions using two, four, and nine person groups. The participant’s task was to assemble pop bead widgets. The group incentive condition resulted in higher performance than did the hourly pay condition.
Allison et al. (1992) studied the job performance of employees working in a treatment center serving handicapped children. They compared performance under the following conditions: hourly pay, individual incentives, cooperative (equally-distributed) group incentives, and competitive (differentially-distributed) group incentives. The group size ranged from 10 – 12 teaching assistants. In the cooperative incentive condition, the amount of incentive earned was based on the average performance of the group, while in the competitive incentive condition the incentive amount was divided between the top three performers in the group. The pay types were superimposed on a feedback system. A within-group reversal design was utilized. Group incentives resulted in higher performance levels than did hourly pay.

Honeywell-Johnson (1997) compared the effects of hourly pay, individual monetary incentives, and equally-distributed group monetary incentives using a within-subject reversal design. Four participants performed a computer task in simulated 10-person groups where they were manipulated into believing they were high performers. Hourly pay resulted in higher performance than did group incentives for three of the four participants. The fourth participant showed steady performance gains over time.

In summary, with one exception, all studies found that group monetary incentives using groups of 2 to 12 members resulted in higher levels of performance than did hourly pay. The results of the Honeywell-Johnson (1997) study appear to be an anomaly and may be due to the use of simulated groups and/or the use of high performers.
Individual Incentives vs. Hourly Pay

All of the studies described above included a comparison of individual incentives with hourly pay (Allison, et al., 1992; Farr, 1976; Honeywell-Johnson, 1997; Roberts & Leary, 1990; Smoot, 1997). Without exception, all found individual incentives to be more effective than hourly pay. These results are consistent with previous research in both laboratory and field settings.

Group Incentives vs. Individual Incentives

Eight studies have compared the effectiveness of group and individual incentives (Allison et al., 1992; Farr, 1976; Honeywell-Johnson, 1997; Honeywell, Dickinson, & Poling, 1997; Roberts & Leary, 1990; Smoot, 1997; Stoneman & Dickinson, 1989; Thurkow, Bailey, & Stamper, 2000). With the exception of Allison et al. (1992) and Thurkow et al. (2000), the studies were conducted in laboratory settings. Farr (1976) found performance levels under individual incentives to be similar to those under the equally-distributed group incentive conditions. However, differentially-distributed group incentives resulted in higher performance levels than individual incentives. Conversely, Allison et al. (1992) found performance under equally-distributed group incentives to be significantly better than individual incentives. No significant differences were found between the differentially-distributed group incentives and individual incentives. In the Roberts and Leary (1990) study, performance levels under individual and group incentives were similar regardless of group size. Similarly, Smoot (1997) found performance comparable between individual and group incentives. Honeywell-Johnson (1997) found that individual incentives resulted in higher performance levels than did group incentives for three of four participants.
Stoneman and Dickinson (1989) compared the effects of individual and equally-distributed group incentives. Groups of two, four, five, or nine participants assembled parts made from nuts, bolts, and washers. An ABA design was utilized with “A” denoting the individual incentive conditions and “B” denoting the group incentive conditions. No performance differences were found between individual and group incentive conditions.

Honeywell et al. (1997) compared the effects of individual and equally-distributed group incentives using an alternating treatments design. Members of the 10 person groups sorted computer data cards with various combinations of holes punched in them. No significant performance differences were found between the individual and group incentive conditions.

Thurkow et al. (2000) used a multi-element design to compare the relative effectiveness of competitive, individual, and group incentive conditions applied to six telephone research interviewers. In the competitive condition, monetary rewards were given to the top producers of each shift. In the individual and group incentive conditions, participants received monetary incentives if performance surpassed a supervisor-set goal. Individual and group incentive conditions both resulted in higher performance than competitive incentives. The greatest productivity rates were achieved with individual incentives.

In summary, with the exception of Honeywell-Johnson (1997) and Thurkow et al. (2000), most research has found group incentives to be at least as effective as individual incentives with groups of 2 to 12 members. Possible reasons the Honeywell-Johnson (1997) results contradicted those from other studies include the focus on high performers, pay differential between conditions, use of simulated work groups resulting in absence of social interaction, type of feedback obtained, and
length of the experimental sessions and/or study. The differences in the results of the Thurkow et al. (2000) study may be explained by the lack of a clear performance standard during the group incentive condition and group size variability.

**Group Size**

As discussed previously, if individual incentives provide a stronger link between performance and pay than do group incentives, it would seem that group incentives should prove less effective than individual incentives at improving performance. Moreover, as group size increases, the link between performance and pay is further weakened and performance should be affected accordingly. Blinder (1990) calls this the 1/n problem. Similarly, Albanese and Van Fleet (1985) discuss the free-rider effect in which a group member, “obtains benefits from group membership but does not bear a proportional share of the costs of providing the benefits.” (p. 244). The authors propose that as group size increases the more likely free-ridership is to occur because in a large group, members are less likely to notice a free-rider’s lack of participation and it is more difficult for a group member to perceive his influence on the group’s output. However, despite a variety of theories that speculate about the detrimental impact of increased group size on performance, research has not borne this out with groups of 2 to 12 members. It may be that small groups allow the individual to maintain sufficient control to reasonably influence group performance and, consequently, the amount of incentive received. That is, the link between performance and pay may be reasonably maintained if group size is not excessive (Honeywell-Johnson & Dickinson, 1999).

There has been little research examining the effects of group size on incentive effectiveness. Two of the previously detailed studies (Stoneman & Dickinson, 1989;
Roberts & Leary, 1990) examined the effects of group size with groups ranging in size from two to nine members. Neither study found differences in performance as a function of group size. No controlled research with larger groups has been conducted. Two field studies, however, have reported decreases in group productivity as group size increased with groups ranging from under 10 to over 50 (Marriott, 1949) and under 20 to over 100 (Campbell, 1952). In both studies, individuals received incentives based on group productivity with the amount prorated based on the worker’s hourly pay and hours worked.

It is unknown what caused the discrepancies between the findings in the laboratory and field settings, however Honeywell-Johnson (1997) speculated that they may be attributed to differences in group size, length of exposure to the pay systems, amount of incentives earned, and/or types of social interactions. While laboratory settings allow researchers to tightly control variables, logistical or monetary concerns often make it difficult to successfully approximate the level of those variables that exist in real world settings. For example, the studies reviewed used groups composed of 2 – 12 members while the field studies had groups of over 100. Using groups of this magnitude would be difficult to accomplish in a laboratory setting. Similarly, laboratory studies are usually conducted over a period of a few days or weeks – a much shorter duration of exposure to an incentive system than is typical for real employees. Similar problems apply in studying the dollar amounts of incentives earned and the types of social interactions that occur in laboratory vs. field settings.

It can be argued that the most adequate method of studying the association between group size and incentive effectiveness requires an examination of the real world. That was the purpose of the current study – to examine the effects of group
size on monetary incentive effectiveness in a real-world setting by conducting a meta analysis of data from incentive systems implemented in 13 companies.
METHOD

Incentive System

The data analyzed in this study were obtained from incentive systems implemented in 13 companies by the Abernathy and Associates consulting firm. These companies, located throughout the United States, were from a variety of industries including banking, manufacturing, sales, trucking, retail, blood banking, and distribution.

The incentive systems utilized a performance scorecard approach (see Table 3 for an example of a performance scorecard). A performance scorecard contains measures determined through a process of cascading objectives. Specifically, the organization’s objectives are defined and then performance measures that align with those organizational objectives are constructed for the next level of the organization. This process is repeated throughout the organization so that measures at each level align with the previous level’s objectives and eventually with the organizational objectives. The cascading objectives approach helps ensure that employees at every level of the organization are focused on achieving the organizational goals.

The scorecard performance scale allows measures with different metrics to be converted to a common scale. In Table 3, a customer service score of 6.5 converts to a 40. The base (B) is set at or near the current performance level and the goal (G) is set at the desired performance level. The weight indicates a measure’s relative importance, with higher weights signifying higher priority measures. The weights of the measures on a scorecard add up to one hundred percent. The weighted score for a particular measure is determined by multiplying the performance scale score by the
weight. The sum of the measures’ weighted scores equals the scorecard performance index.

Table 3
Performance Scorecard Example

<table>
<thead>
<tr>
<th>Measure</th>
<th>-10</th>
<th>0</th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>80</th>
<th>100</th>
<th>Weight</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>percent deadlines met</td>
<td>0</td>
<td>10</td>
<td>20</td>
<td>30</td>
<td>40</td>
<td>50</td>
<td>60</td>
<td>70</td>
<td>90</td>
<td>100</td>
<td>.10</td>
<td>2</td>
</tr>
<tr>
<td>Customer Service Score</td>
<td>4.0</td>
<td>4.5</td>
<td>5.0</td>
<td>5.5</td>
<td>6.0</td>
<td>6.5</td>
<td>7.0</td>
<td>7.5</td>
<td>8.0</td>
<td>8.5</td>
<td>.30</td>
<td>12</td>
</tr>
<tr>
<td>Gross Revenue (000)</td>
<td>10</td>
<td>15</td>
<td>20</td>
<td>25</td>
<td>30</td>
<td>35</td>
<td>40</td>
<td>45</td>
<td>50</td>
<td>55</td>
<td>.60</td>
<td>48</td>
</tr>
<tr>
<td>base</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>goal</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The amount of incentive pay available (opportunity) is either a fixed amount (budgeted), linked to gains in organizational or divisional net income (performance-indexed), or a combination of both. The amount paid to employees is calculated by multiplying the opportunity by the scorecard performance index. Interested readers can refer to William B. Abernathy’s *The Sin of Wages* (1996) and *How to Design Effective Incentive Plans* (1993) for a detailed discussion of the Performance Scorecard approach. Similar methodologies of organizational and performance measurement include Kaplan and Norton’s “Balanced Scorecard” and Daniels “Performance Matrix”.
Variables

Dependent Variables

This study examined the actual measurement scores as opposed to the weighted performance scores. Scores were obtained once per month; thus a measure having nine months of data had nine scores. On the advice of Abernathy and Associates, each measure’s first three data points were omitted. This was suggested because it typically took about three months for the incentive system to be fully implemented. During this initial implementation period, persons in the organizations were learning to obtain accurate measurements and issues in the computerized reporting system were being resolved. These two conditions caused the reliability of the initial data to be suspect and thus they were omitted from this study. In addition, measures with less than six months of data were not included in this study.

Independent Variables

Group Size

A measure’s group size was calculated by summing the number of people with that measure on their scorecard. Group size is not necessarily the number of people in an organizational unit (e.g., team or department) but the number of people on which performance was assessed by a particular measure. This was sometimes a subset of a department or a combination of multiple departments. For example, if a person’s scorecard contained five measures they could consist of two measures that assessed her individual performance (group size = 1); two that assessed her team’s performance (e.g., group size = 5); and one that assessed her department’s performance (e.g., group size = 35). When these incentive systems were
implemented, managers received training in how to explain the scorecard system to their employees. In addition, in some companies, employees also received an orientation from Abernathy and Associates consultants explaining the incentive system. Through these activities, employees had a good understanding of whether the measures on their scorecards were assessing individual or group performance as well as the group size for each measure. Occasionally, a supervisor’s scorecard included the measures of their direct reports. The supervisor was not included in the calculation of group size because his or her performance was not measured.

Measure Level

Each measure was coded according to the organizational level targeted for performance improvement by that measure. Measure level is not the level of the people whose performance is assessed by that measure, but the level at which the measure is evaluated. For example, a team may have the measure “customer-service” on their scorecard that is measured at the division level. This “customer-service” measure would be coded as a divisional measure, not a team measure. Measure levels included organizational, division, department, team, individual, and undefined.

Long/Short-Term Profitability

Each measure was coded according to how it related to the profitability of the company (i.e., long-term, short-term, and other). Long-term profitability measures were those designed to improve customer service, regulatory compliance, and project performance. Short-term profitability measures were those designed to improve sales,
productivity, and expense control. A coding of “other” indicated either an administrative checklist measure or a non-defined measure.

**Length of Exposure**

The length of exposure to the incentive system was defined as the number of months of data for that measure. The length of exposure did not include the omitted first three data points.

**Company**

Each measure was coded to associate it with company from which it originated. This allowed the researcher to make cross-company comparisons on variables that may have affected the results. Known company variables included, (a) the existence of Performance Leadership Training, (b) the incentive pay type (Budgeted, Profit-Indexed, or Mixed), and (c) incentive pay frequency (monthly or quarterly). See Table 4 for a listing of known company variables.
Table 4

Cross-Classification of Company Variables and Company Code

<table>
<thead>
<tr>
<th>Company Variables</th>
<th>Company Code</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A  B  C  D  E  F  G  H  I  J  K  L  M</td>
</tr>
<tr>
<td>(a) Training</td>
<td>Y  N  N  N  N  N  N  Y  Y  N  N  N  N</td>
</tr>
<tr>
<td>(b) Pay Type</td>
<td>P  P  M  P  M  B  P  P  P  P  P  P  B</td>
</tr>
<tr>
<td>(c) Pay Frequency</td>
<td>M  Q  M  M  M  M  M  M  M  M  M  M  M</td>
</tr>
</tbody>
</table>

Legend.  
Training:  
Y = Yes  
N = No  
Pay Type:  
P = Performance Indexed  
B = Budgeted  
M = Mixed  
Pay Frequency:  
M = Monthly Payout  
Q = Quarterly Payout

Types of Comparisons

Three types of comparisons were defined:

1. Between-Group – Within-Company. These comparisons were based on a measure that was in place in two or more groups within an organization as shown in Figure 2.
Figure 2. Between-Group – Within-Company Comparison Example.

2. Between-Group – Across-Company. These comparisons were based on a measure that existed in two or more companies as shown in Figure 3.
3. Group-Expansion. These comparisons were based on a measure that was in place in one unit of the organization and the group size increased at some point during the observation period as shown in Figure 4.

![Number of Widgets Produced per Month](chart.png)

Figure 4. Group-Expansion Comparison Example.

Analyses

**Between-Group – Within-Company Comparisons of the Time/Performance Relationship**

The correlation between the month (e.g., 1, 2, 3…) of measurement and the performance score achieved was calculated for each Between-Group – Within-Company measure. This correlation coefficient provides a standardized measure of performance increase that allows comparisons across measures having different metrics (e.g., dollars, percentages, and averages). This correlation was transformed to
Fisher’s Z to facilitate subsequent statistical analyses. Analyses of the obtained Z’s were classified into three group size categories (i.e., Individual = 1 member; Small = 2 – 12 members; Large = 13 or more members).

The average weighted Fisher’s Z was calculated for each group size category using the formula:

\[
\frac{\sum_{j=1}^{J} (n_j - 3)Z_j}{\sum_{j=1}^{J} (n_j - 3)} = Z
\]

The average Z was then transformed to obtain the average correlation for each group size category.

A one-way analysis of variance (ANOVA) was conducted to test the hypothesis that the means of the Fisher’s Z’s for the three group sizes were equal. An alternative test, likely to be more sensitive (Bradley Huitema, personal communication, March 20, 2001), was also conducted. The latter test is a modification of Fisher’s X² test for the homogeneity of a set of correlation coefficients. This overall test for differences among the three groups in terms of average Fisher’s Z used the formula:

\[
\sum_{j=1}^{3} (\sum n_i - 3)^2 Z_j = \frac{\left[ \sum_{j=1}^{3} (\sum n_j - 3)Z_j \right]^2}{\sum_{j=1}^{3} (\sum n_j - 3)} = X_{\text{ow}}^2
\]

where \( n_j \) was the sample size, \( Z_j \) was the value of Fisher’s Z transformed from the jth correlation coefficient, \( J \) was the number of samples, and \( X_{\text{ow}}^2 \) was the obtained chi-squared test statistic. Results were compared against \( X_{J-1}^2 \) with alpha set at 0.05.
A test for differences between two average Fisher’s Z’s was conducted to determine where significant difference(s) existed (with alpha set at 0.05) using the formula:

\[
\frac{\bar{Z}_1 - \bar{Z}_\mu}{\sqrt{\frac{1}{\sum_{i=1}^{n} (n_j - 3)_i} + \frac{1}{\sum_{i=1}^{n} (n_j - 3)_\mu}}} = z
\]

where \(z\) is evaluated as a normal deviate.

**Time/Performance Relationship within Group Size Categories**

Fisher’s (1958) test of the homogeneity of a set of sample correlation coefficients was conducted to determine if the variation among correlations computed between time and performance within each group size category was explainable by sampling error.

**Association with Predictors**

Regression and correlation analyses to determine the contribution of the predictors separately and in combination in predicting Fisher’s Z were conducted within each group size category. In addition, regression and part correlation analyses were conducted to determine the unique contribution of a sub-set of predictors above and beyond the influence of the other predictors. The latter analyses included the part correlation \(r_y(x_i \cdot x_1, x_2, \ldots x_{m-i})\) where

\[
r_y(x_i \cdot x_1, x_2, \ldots x_{m-i}) = \sqrt{\frac{SS_{uniquepred \ x_i}}{SS_{Total}}},
\]

and the following F-test on this correlation
\[
\frac{SS_{\text{unique pred}}}{df_{\text{reg (all)}} - df_{\text{reg (reduced)}}} = F,
\]

where \(SS_{\text{unique-xi}}\) is the difference between the full model (all predictors) regression sum of squares and the reduced model regression sum of squares that excludes variable \(x_i\) as a predictor, \(df_{\text{reg (all)}}\) is the regression degrees of freedom for the full model, \(df_{\text{reg (reduced)}}\) is the regression degrees of freedom for the reduced model, and \(MS_{\text{resid (all)}}\) is the residual mean square for the full model. Alpha was set at 0.05.

**Between-Group – Across-Company Comparisons of the Time/Performance Relationship**

The same analyses described above for the Between-Group – Within-Company comparison design were applied to the Between-Group – Across-Company data, with one exception. Correlation and regression analyses were not computed within groups to measure the association between performance and organizational predictors because sample size was too small.

**Group-Expansion Comparisons**

Each Group-Expansion data set was analyzed using two time-series regression models. These models and the criterion for selecting the better of the two are described next.

**Two-Parameter Model**

Data from each Group-Expansion comparison were analyzed using the two-parameter time-series regression model \(Y_t = \beta_0 + \beta_1 X_t + \epsilon_t\). The model was fit using the design matrix shown in Figure 5. To determine the mean difference between the
pre- and post-intervention phases, regressions were conducted on each measure where
$Y_t$ was the performance at time $t$ and $X$ was a dummy variable indicating whether the
respective performance occurred before or after the group size increased (0 and 1,
respectively). The standardized effect was obtained for each comparison to determine
how far the phase means differed from each other in standard deviation units thus
allowing analyses of comparisons having different metrics. The standardized effect
was obtained by dividing the estimated mean difference, $b_1$, by the estimate of the
standard error of estimate, $s$. The sign was reversed when the measure was of the type
where a decrease in that measure was an improvement in performance.
The lack of autocorrelation (i.e., independence of residuals) is assumed in linear regression. The lag-1 autocorrelation coefficient was obtained for each estimated Group-Expansion comparison; the H-M test (Huitema & McKean, 2000) was then computed to evaluate the existence of autocorrelated errors. This test is based on the following test statistic:

\[
\frac{r_1 + \frac{P}{N}}{\sqrt{\frac{(N - 2)^2}{(N - 1)N^2}}} = z_{HM}
\]
where \( r_1 \) denotes the autocorrelation coefficient, \( P \) denotes the number of parameters in the time-series regression model, \( N \) denotes the total number of observations in the time-series, and \( z_{hm} \) is evaluated as a standard normal deviate.

**Four-Parameter Model**

Data from each Group-Expansion comparison were analyzed using the four-parameter time-series regression model \( Y_t = \beta_0 + \beta_1 t + \beta_2 D + \beta_3 SC + \varepsilon_t \). The model was estimated using the Huitema & McKean (2000) design matrix shown in Figure 6. To determine the size of the change in level between the pre- and post-intervention phases, regressions were conducted on each measure where \( Y_t \) was the performance at time \( t \) and the three predictor variables \( t, D, \) and \( SC \) were used. The first predictor, \( t \) was the observation number (1, 2, 3, etc.). The second predictor \( D \) was a dummy variable indicating whether the respective performance occurred before or after the group size increased (0 and 1, respectively). The third predictor was \( SC \), the slope change variable (0 for observations occurring before the group size change and the first observation after the group size change; 1, 2, 3, etc. for the observations thereafter). The standardized effect was obtained for each comparison to determine the level change in standard deviation units. This allowed analyses of comparisons having different metrics. The standardized effect was obtained by dividing the pre/post group size change coefficient \( b_2 \) by the estimate of the standard error of estimate. As with the two-parameter model, the lag-1 autocorrelation coefficient and \( z_{hm} \) were calculated for each set of Group-Expansion observations.
Legend. \( n_1 \) = the number of observations in the pre-intervention phase
\( n_2 \) = the number of observations in the post-intervention phase

Figure 6. Matrices for Four-parameter Two-phase Intervention Model (Huijtema and McKean, 2000).

Model Selection

After the two- and four-parameter models were estimated, a choice between these models was made on the basis of the estimated standard error of estimate. The model with the smaller standard error was selected as the preferred model for describing the data. Subsequent analyses of average effect size and average autocorrelation were based on the results obtained from the preferred model.
Tests for Effects of Increase in Group Size

The question of whether there was any effect on performance when the group size changed was answered using one-sample t and Wilcoxon signed rank tests. Additionally, each comparison was labeled by type of group size change. Comparisons where the group size changed from an individual to a group (i.e., an increase from one member to two members) were labeled as “One.” When the initial group size was 2 or higher and increased by one or two members with it was labeled as a “Small” group size change. Lastly, when the group size changed from two to seven members it was labeled as a “Large” change. One-way ANOVA and Kruskal-Wallis tests were conducted to examine the effect of change type (One, Small, or Large) on the means and medians, respectively, of the standardized effect estimates.
RESULTS

Between-Group – Within-Company Comparisons

There were 513 Between-Group – Within-Company data sets (see Appendix A for Between-Group – Within-Company comparison graphs). These were categorized into the number of data sets available in each group size type. Table 5 shows the number of data sets associated with each level of the four predictor variables: Measure Level, Long-Term/Short-Term Profitability Measure Type, Company, and Exposure (e.g., number of months of data collection) by group size type. For example, of the 90 Individual group size data sets, there was one Division level measure; 13 Department level measures; 4 team level measures; 69 Individual level measures; and 3 undefined.

Table 5

Number of Data Sets Associated with each Level of the Four Predictor Variables (Between-Group – Within-Company Comparisons)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Predictor Levels</th>
<th>Individual</th>
<th>Small</th>
<th>Large</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Division</td>
<td>1</td>
<td>11</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>Level</td>
<td>Department</td>
<td>13</td>
<td>139</td>
<td>22</td>
<td>174</td>
</tr>
<tr>
<td></td>
<td>Team</td>
<td>4</td>
<td>39</td>
<td>2</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>Individual</td>
<td>69</td>
<td>182</td>
<td>21</td>
<td>272</td>
</tr>
<tr>
<td></td>
<td>Undefined</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>9</td>
</tr>
</tbody>
</table>

(Table 5 continues)
Table 5 – Continued

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Predictor Levels</th>
<th>Individual</th>
<th>Small</th>
<th>Large</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Profitability</td>
<td>Long</td>
<td>9</td>
<td>56</td>
<td>10</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td>Short</td>
<td>45</td>
<td>233</td>
<td>29</td>
<td>307</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>36</td>
<td>88</td>
<td>7</td>
<td>131</td>
</tr>
<tr>
<td>Company</td>
<td>A</td>
<td>3</td>
<td>18</td>
<td>-</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>8</td>
<td>58</td>
<td>-</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>7</td>
<td>20</td>
<td>-</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>-</td>
<td>-</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>52</td>
<td>158</td>
<td>3</td>
<td>213</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>9</td>
<td>19</td>
<td>1</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>H</td>
<td>5</td>
<td>25</td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>2</td>
<td>22</td>
<td>19</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>1</td>
<td>9</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>1</td>
<td>19</td>
<td>-</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>1</td>
<td>28</td>
<td>1</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>1</td>
<td>1</td>
<td>-</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exposure (number of months of data collection)</th>
<th>Range (mean)</th>
<th>6 - 67</th>
<th>7 - 67</th>
<th>6 - 37</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(mean)</td>
<td>(16.19)</td>
<td>(16.20)</td>
<td>(20.65)</td>
</tr>
</tbody>
</table>

| Total | 90    | 377    | 46     | 513    |
Between-Group – Within-Company Differences in Time/Performance Relationship

Figure 7 displays the average weighted Fisher’s Z for each group size type. Average weighted Fisher’s Zs for the Individual, Small, and Large groups were 0.21, 0.17, and 0.01, respectively. The corresponding average correlations were 0.21, 0.17, and 0.01. As shown, there was a very small difference between the Individual and Small group average weighted Fisher’s Zs and a relatively large difference between those two groups and the Large Group. Overall, the average weighted Fisher’s Z appears to decrease as group size increases.

![Figure 7](image_url)

Figure 7. Average Weighted Fisher’s Z for each Group Size (Between-Group – Within-Company).

A one-way ANOVA was conducted on the average Fisher’s Zs. The differences between the three groups were not statistically significant ($F = 2.26$, $df = 512$, $p = 0.11$). Table 6 displays the source table.
Table 6
Source Table for the ANOVA Conducted on the Average Fisher’s Z for the Individual, Small, and Large Groups (Between-Group – Within-Company)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2</td>
<td>1.12</td>
<td>0.56</td>
<td>2.26*</td>
</tr>
<tr>
<td>Error</td>
<td>510</td>
<td>126.10</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>512</td>
<td>127.22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p = 0.11

The modified Fisher $X^2$ test, however, yielded a statistically significant result ($p = 0.00$). The modified Fisher tests on differences between two groups found that there was no significant difference between the Individual and Small groups ($p = 0.10$) but there were significant differences between the Individual and Large groups ($p = 0.00$) and the Small and Large groups ($p = 0.00$). In short, the Individual and Small groups’ average Fisher’s Zs were not significantly different from each other but they were significantly different from the Large group.

Time/Performance Relationship within Group Size Categories

Highly significant differences among the Z’s were found within the Individual ($X^2 = 167.92$, df = 89, $p = 0.00$), Small ($X^2 = 1244.37$, df = 376, $p = 0.00$), and Large groups ($X^2 = 363.71$, df = 45, $p = 0.00$). Hence, differences in the time/performance relationship appeared to be systematic within all group size categories. That is, variation among the Z’s within the groups was not explainable by sampling error.
alone. This outcome suggests that there may be variables that predict some of this systematic variation.

**Association with Predictors**

Regression analyses were conducted for each group size separately with Fisher’s Z as the response and each of the four predictors (separately and in combination (All)) as the predictors. Table 7 displays the obtained multiple correlations and p-values. The only significant result found within the Individual group was with all of the predictors in combination (p = 0.03). The separate predictor analyses produced no statistically significant correlations in the Individual group. With the exception of Measure Level (p = 0.06), significant results were found in the Small group for each of the predictors separately and in combination. None of the predictors analyzed separately or in combination resulted in significant findings for the Large group.
Table 7
Multiple correlation coefficients (and Obtained p Values) for All Predictors in Each Group Size (Between-Group – Within-Company)

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Group</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Individual</td>
<td>Small</td>
<td>Large</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(n = 90)</td>
<td>(n = 377)</td>
<td>(n = 46)</td>
<td></td>
</tr>
<tr>
<td>Company</td>
<td>0.40 (0.16)</td>
<td>0.30 (0.00)</td>
<td>0.17 (0.94)</td>
<td></td>
</tr>
<tr>
<td>Exposure</td>
<td>0.07 (0.52)</td>
<td>0.22 (0.00)</td>
<td>0.12 (0.43)</td>
<td></td>
</tr>
<tr>
<td>Measure Level</td>
<td>0.18 (0.58)</td>
<td>0.16 (0.06)</td>
<td>0.24 (0.48)</td>
<td></td>
</tr>
<tr>
<td>Long/Short-term</td>
<td>0.14 (0.43)</td>
<td>0.15 (0.01)</td>
<td>0.22 (0.36)</td>
<td></td>
</tr>
<tr>
<td>Profitability</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All Predictors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined</td>
<td>0.54 (0.03)</td>
<td>0.39 (0.00)</td>
<td>0.37 (0.75)</td>
<td></td>
</tr>
</tbody>
</table>

Part correlation analyses were conducted to determine the unique contribution of the predictors above and beyond the influence of the other predictors. Analyses were done for each group size separately. Then the data from all groups were combined and analyzed as a single combined group. Table 8 displays the summary of the obtained part correlations and p-values. Overall, the results were as follows:

1. Company. Significant results were found in the Individual and Small groups and all groups in combination. A significant result was not found in the Large group.
2. Exposure. The only significant result was found in the Small group.
3. Measure Level. No significant results were found
4. Long/Short-term profitability. Significant results were found in the Individual and Small groups.

5. Group size type. A significant result was found when examining all groups in combination.

Table 8

Obtained Part Correlation Coefficients (and p Values) from Part Correlation Analyses (Between-Group – Within-Company)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Individual (n = 90)</th>
<th>Small (n = 377)</th>
<th>Large (n = 46)</th>
<th>All Groups Combined</th>
</tr>
</thead>
<tbody>
<tr>
<td>Company</td>
<td>.46 (.02)</td>
<td>.68 (.00)</td>
<td>.26 (.53)</td>
<td>.20 (.02)</td>
</tr>
<tr>
<td>Exposure</td>
<td>.11 (.28)</td>
<td>.44 (.00)</td>
<td>.11 (.54)</td>
<td>.12 (.32)</td>
</tr>
<tr>
<td>Measure Level</td>
<td>.12 (.66)</td>
<td>.32 (.22)</td>
<td>.29 (.11)</td>
<td>.11 (.13)</td>
</tr>
<tr>
<td>Long/Short-term profitability</td>
<td>.34 (.00)</td>
<td>.35 (.03)</td>
<td>.09 (.87)</td>
<td>.11 (.39)</td>
</tr>
<tr>
<td>Group Size Type</td>
<td>(n/a)</td>
<td>(n/a)</td>
<td>(n/a)</td>
<td>.12 (.00)</td>
</tr>
</tbody>
</table>

Between-Group – Across-Company Comparisons

There were 41 Between-Group – Across-Company measures (see Appendix B for Between-Group – Across-Company comparison graphs). These were categorized into the Individual (n = 12), Small (n = 27), and Large (n = 2) group categories. Table 9 shows the breakdown of the predictors Measure Level, Long-Term/Short-Term Profitability, Measure Type, Company, and Exposure (e.g., number of months of data collection) by group size type.
Table 9
Number of Data Sets Associated with each Level of the Four Predictor Variables (Between-Group – Across-Company Comparisons)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Predictor Levels</th>
<th>Individual</th>
<th>Small</th>
<th>Large</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measure</td>
<td>Organizational</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Level</td>
<td>Division</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Department</td>
<td>4</td>
<td>17</td>
<td>1</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Team</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Individual</td>
<td>7</td>
<td>7</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>Company</td>
<td>A</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>G</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>I</td>
<td>0</td>
<td>4</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>J</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Exposure</td>
<td>Range</td>
<td>9 – 67</td>
<td>7 – 67</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>(number of months of data collection)</td>
<td>(mean)</td>
<td>(23.50)</td>
<td>(24.19)</td>
<td>(27.00)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>12</td>
<td>27</td>
<td>2</td>
<td>41</td>
</tr>
</tbody>
</table>
Figure 8 displays the average weighted Fisher’s Z for each group size type. Average weighted Fisher’s Zs for the Individual, Small, and Large groups were 0.04, 0.13, and 0.18, respectively. The corresponding average correlation coefficients were 0.04, 0.13, and 0.18. As shown, there were small differences between each of the groups; the average weighted Fisher’s Z appeared to increase as group size increased.

A one-way ANOVA was conducted on the average Fisher’s Zs. The differences between the three groups were not statistically significant (F = 0.41, df = 40, p = 0.67). Table 10 displays the source table.
Table 10

Source Table for the ANOVA Conducted on the Average Fisher’s Z for the Individual, Small, and Large Groups (Between-Group – Across-Company Comparisons)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
<td>2</td>
<td>0.22</td>
<td>0.11</td>
<td>0.41</td>
</tr>
<tr>
<td>Error</td>
<td>38</td>
<td>10.26</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>10.48</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p = 0.67

The results of the modified Fisher test (p = 0.42) were consistent with those of the ANOVA.

Time/Performance Relationship Within Group Size Categories

Highly significant results were found within the Individual ($X^2 = 25.03$, df = 11, $p = 0.01$), Small ($X^2 = 160.57$, df = 26, $p = 0.00$), and Large groups ($X^2 = 4.94$, df = 1, $p = 0.03$). Hence, for all groups, differences in the time/performance relationship appeared to be systematic. That is, variation within the groups was not explainable by sampling error alone. Subsequent analyses of possible relationships between variation in performance within groups and organizational predictor variables were not performed because sample size was small (n = 41).
Group-Expansion Comparisons

There were 47 Group-Expansion comparisons (see Appendix C for Group-Expansion comparison graphs). The mean and median standardized effects were –0.42 and –0.42, respectively. Figure 9 shows the distribution of the standardized effects for all 47 data sets. While the one-sample $t$ test did not quite reach statistical significance ($p = 0.06$), the more robust Wilcoxon Signed Rank Test did show a statistically significant effect of the change in group size ($p = 0.01$). Hence it is concluded that there was a moderate decrease in performance when group size increased.

-4.8 -3.2 -1.6 0.0 1.6 3.2

Figure 9. Dotplot of the Standardized Effects.

There were 20, 19, and 8 data sets, respectively, within the One, Small, and Large size change groups. A one-way ANOVA was conducted on the mean standardized effects. Table 11 displays the source table. The mean differences among the three groups were not statistically significant ($F = 0.06$, df = 46, $p = 0.94$). There does, however, appear to be a relationship between the change size ratio (the group size after the change in group size divided by the group size prior to the group size change) and the standardized effect mean for that group size change type. That is, as shown in Figure 10, the mean standardized effect appears to decrease as the change size ratio increases. The degree of this relationship, however, is not statistically significant.
Table 11
Source Table for the ANOVA Conducted on the Standardized Effects for the One, Small, and Large Change Type Groups (Group-Expansion Comparisons)

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change Type</td>
<td>2</td>
<td>0.27</td>
<td>0.13</td>
<td>0.06*</td>
</tr>
<tr>
<td>Error</td>
<td>44</td>
<td>101.72</td>
<td>2.31</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>46</td>
<td>101.99</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p = 0.94

Figure 10. Mean Standardized Effect by Change Ratio (Group-Expansion Comparisons).
The Kruskal-Wallis test did not result in any significant differences between the change type groups in terms of median standardized effect (p = 0.77). See Table 12. However, as with the ANOVA, a relationship between the ratio of change size and the standardized effect median for that group size change type is somewhat suggestive (see Figure 11).

Table 12

Results Table for the Kruskal-Wallis Conducted on the Standardized Effects for the One, Small, and Large Change Type Groups (Group-Expansion Comparisons)

<table>
<thead>
<tr>
<th>Change Type</th>
<th>N</th>
<th>Median</th>
<th>Ave Rank</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>One</td>
<td>20</td>
<td>-0.51</td>
<td>22.70</td>
<td>-0.56</td>
</tr>
<tr>
<td>Small</td>
<td>19</td>
<td>-0.17</td>
<td>25.70</td>
<td>0.72</td>
</tr>
<tr>
<td>Large</td>
<td>08</td>
<td>-0.41</td>
<td>23.10</td>
<td>-0.20</td>
</tr>
<tr>
<td>Overall</td>
<td>47</td>
<td></td>
<td>24.00</td>
<td></td>
</tr>
</tbody>
</table>

h = 0.52  df = 2  p = 0.77
The distributions of the lag-1 autocorrelation coefficient and the test statistic \( Z_{hm} \) are displayed in Figures 12 and 13, respectively. The degree of autocorrelation is approximately what would be expected by chance, thus, autocorrelation did not have a major influence in estimating change.

![Figure 11. Median Standardized Effect by Change Ratio (Group-Expansion Comparisons).](image)

![Figure 12. Dotplot of Group-Expansion Analyses Lag-1 Autocorrelation Coefficients.](image)

![Figure 13. Dotplot of Group-Expansion Analyses \( Z_{hm} \).](image)
DISCUSSION

If individual incentives provide a stronger link between performance and pay than do group incentives, individual incentives should prove more effective than group incentives at improving performance. Moreover, as group size increases, the diminished link between performance and pay should weaken performance. The few controlled studies that have examined the effects of group size on incentive effectiveness have not found differences in performance as a function of group size with groups containing two through nine members. No controlled research with larger groups has been conducted, but two field studies reported decreases in group productivity as group size increased with groups ranging from under ten to over one hundred. Honeywell-Johnson (1997) speculated that the discrepancies between the laboratory and field studies might be due to differences in group size, length of exposure to the pay systems, amount of incentives earned, and/or types of social interactions. Given that it is often difficult to successfully approximate real world levels of those variables in laboratory settings, the current study was designed to examine the effects of group size on incentive effectiveness by using data from incentive systems implemented in applied settings.

In this study the most robust analysis of the data based on increasing the group size (Group-Expansion) showed significant changes in performance associated with an increase in group size. However, other analyses showed no statistically significant association between performance and the size of the group size change (i.e., One, Small, and Large). These latter analyses did, however, suggest a weak relationship between the change size ratio and performance. These results are consistent with the
previous laboratory research that found no significant differences between the
effectiveness of individual and small group incentives.

Similarly, differences in group size were not associated with performance
differences in the Between-Group – Across-Company analyses. However, there were
only two data sets with group sizes over 12 members. This restriction of range may
have influenced the results. Future researchers should seek organizations having
more variation among group sizes. In addition, because the data were compared
across different companies, there undoubtedly were unknown organizational
variable(s) that affected the results. Future researchers should consider
simultaneously measuring other organizational variables to determine their possible
influence on the results.

Unlike the Between-Group – Across-Company comparisons, the Between-
Group – Within-Company comparisons had an adequate amount of data within the
different group size categories from which to conduct analyses. The results of the
analyses on the Between-Group – Within-Company comparison were consistent with
the literature. Overall, performance appeared to decrease as group size increased.
The tests for differences confirmed the existence of significant effects of different
group sizes. When the data were examined more closely, it was determined that
performance in the Individual and Small groups was not significantly different but
both groups were significantly different from the Large group. Data were not
available on all of the variables hypothesized by Honeywell-Johnson (1997) to
influence results. However, information on the group size and length of exposure to
the pay systems was available. Group size type was found to be a significant
predictor. Length of exposure to the incentive system was only found to be a
significant predictor for the Small group size group. The outcome associated with the other predictors was inconsistent.

The results of this study are consistent with previous research and theory regarding the effects of group size on incentive effectiveness. That is, individual and small group incentives appear to be equally effective but this effectiveness decreases as the group size increases. As Smoot (1997) described, “…the power of the pay-for-performance contingency to motivate individual workers to work more and work better is dependent, in part, upon the extent to which individual workers can control pay levels as a function of their own behavior.” (p. 3). With individual incentives, the pay level is directly contingent upon the performance of an individual worker whereas with group incentives, an individual’s level of pay is dependent upon his or her own performance as well as that of the other members of the incentive group. As the group size increases, the pay level becomes less and less contingent upon the performance of that individual worker and, consequently, more out of his or her direct control. However, this and other studies have found individual and small group incentives (2 - 12 workers) to be equally effective. It may be that with small group incentives an individual worker is able to maintain a sufficient amount of control over the group’s performance and, as a result, the level of incentive pay received. This may be particularly true with high performers who are “carrying” the group in which a decrease in their performance would negatively impact the level of pay received. In a large group, the ability of an individual worker to influence the overall performance of the group and the amount of incentive pay earned is decreased. That is, the extent to which the pay level for an individual worker is contingent upon the performance of that worker is increasingly diminished as the group size increases and, thus, performance should decrease. In addition, in a small group, both high and low
performing workers can be easily identified by others in their work group and subjected to the associated social reinforcers and punishers from other workers. In large groups, the contribution of any particular worker is more difficult for other group members to identify and, consequently, reward or punish. Finally, it is easier for workers in both individual and small groups to determine current performance levels and, thus, estimate the amount of incentive to be earned. This would get increasingly more difficult as the number of group members increased and may influence the effectiveness of the incentive system.

This study is unique in that it is the first known study that has examined the effects of group size on incentive systems implemented in multiple organizations. The triangulation of the three different comparisons (Between-Group – Within-Company, Between-Group – Across Company, and Group Expansion) allowed a study of the research question from different angles, thus highlighting any consistency of effects across the comparison types. In addition, because these analyses were conducted from incentive systems implemented in real companies – as opposed to contrived laboratory settings – the results obtained will be more likely to generalize to other organizations. Obtaining such realism, however, comes at the expense of experimental control. It is possible that there were important unrecognized variables affecting the results. Thus, future researchers should consider utilizing more sophisticated designs that incorporate measurements such as the length of exposure to the pay systems, amount of incentives earned, and types of social interactions. Effort should be made to obtain large data sets that will allow adequate examination of these and other important organizational variables. Moreover, the effects of group size aside, this study assumed that the incentive systems implemented had an overall positive impact on
performance. However, without access to baseline data we cannot say with certainty that the incentive systems improved performance. This was further impacted by the necessity of deleting the first three months of data when it can be assumed that any performance improvements would be most evident. Future research would benefit from improvements in the instrumentation used to measure performance and/or the development of methods to allow those collecting measurements to achieve a level of sophistication more quickly. Either of these enhancements would give researchers the ability to analyze data immediately after the incentive systems were implemented when it can be assumed that the impact of an effective incentive system would be most noticeable. Better still would be the gathering of baseline data prior to the incentive systems implementation. Finally, given the small volume of research examining the effects of group size on incentive effectiveness, future research is needed to more adequately evaluate the controlling variables.
Appendix A

Between-Group – Within-Company Comparison Graphs
### Ad Placement Acc-%-min

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Average Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>-0.50</td>
</tr>
<tr>
<td>4</td>
<td>-0.40</td>
</tr>
<tr>
<td>5</td>
<td>-0.30</td>
</tr>
<tr>
<td>6</td>
<td>-0.20</td>
</tr>
<tr>
<td>7</td>
<td>-0.10</td>
</tr>
<tr>
<td>8</td>
<td>0.00</td>
</tr>
<tr>
<td>9</td>
<td>0.10</td>
</tr>
<tr>
<td>10</td>
<td>0.20</td>
</tr>
</tbody>
</table>

### %Onetime Projects

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Average Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.70</td>
</tr>
<tr>
<td>3</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Accidents Cost/Mile - NEG

Group Size

Average Correlation

Non-TCW Mtnce / Mile - NEG

Group Size

Average Correlation
% No Errors Reported

Average Correlation

Group Size

Cust Impact Errors - NEG

Average Correlation

Group Size
Company Packages

Group Size  Average Correlation

1  0.05
2  0.40

Cust&Bank ImpctErrRt - NEG

Group Size  Average Correlation

8  0.00
9  -0.10
Functional Errors

Group Size	Average Correlation

Individual Error Rte - NEG

Group Size	Average Correlation
**Items Processed/Hour**

- Group Size: 2
- Average Correlation: 0.40

- Group Size: 6
- Average Correlation: 0.45

- Group Size: 7
- Average Correlation: 0.15

- Group Size: 8
- Average Correlation: 0.20

- Group Size: 9
- Average Correlation: 0.30

**Keystrokes per Hour**

- Group Size: 2
- Average Correlation: 0.30

- Group Size: 6
- Average Correlation: 0.40
Perfect Attendance

Group Size

Average Correlation

0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00
1.10

1 2 3 4 5 6 7 8 9 10 11 12 13

# Photos / Hour (Co. 35)

Group Size

Average Correlation

0.00
0.10
0.20
0.30
0.40
0.50
0.60

4 6
### LTL Stops

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Average Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>-0.40</td>
</tr>
<tr>
<td>10</td>
<td>-0.30</td>
</tr>
<tr>
<td>19</td>
<td>-0.20</td>
</tr>
<tr>
<td>53</td>
<td>-0.10</td>
</tr>
</tbody>
</table>

### TL Stops

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Average Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>0.00</td>
</tr>
<tr>
<td>10</td>
<td>0.20</td>
</tr>
<tr>
<td>19</td>
<td>0.40</td>
</tr>
<tr>
<td>53</td>
<td>0.60</td>
</tr>
</tbody>
</table>
Miles Driven

Average Correlation

% On Time & Acc. PU & Delvry

Average Correlation

Group Size

Group Size
SOP Turnaround Time - NEG

% Deferrals + % QNS - NEG
Last Audit Score

![Bar Chart](chart1.png)

Group Size

Drive Chair Satisfy

![Bar Chart](chart2.png)

Group Size
% Orders Returned - NEG

![Graph showing % Orders Returned - NEG](image)

Average Correlation

Inventory Turnover

![Graph showing Inventory Turnover](image)
### Safety/ Housekeeping

![Bar chart showing average correlation for different group sizes.]

### Revenue per Hour

![Bar chart showing average correlation for different group sizes.]

- The chart on the left illustrates the relationship between group size and average correlation in the Safety/Housekeeping category. The bars indicate an increase in correlation as group size increases from 5 to 10.

- The chart on the right shows the relationship for Revenue per Hour. Similar to the previous chart, there is an observed increase in correlation with increasing group size, from 2 to 10.
Appendix B

Between-Group – Across-Company Comparison Graphs
DSO - NEG

Group Size

Average Correlation

Cost/ Position Filled - NEG

Group Size
### %OnTime Reports

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Average Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.10</td>
</tr>
<tr>
<td>3</td>
<td>-0.30</td>
</tr>
<tr>
<td>6</td>
<td>-0.40</td>
</tr>
</tbody>
</table>

### Avg Days to Fill Pos - NEG

<table>
<thead>
<tr>
<th>Group Size</th>
<th>Average Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.07</td>
</tr>
<tr>
<td>8</td>
<td>0.07</td>
</tr>
</tbody>
</table>
Payroll/Revenue - NEG

Average Correlation

% Days Balanced

Average Correlation

Group Size 1 2 3 4 6 7 9 19

Group Size 1 2 4 6
% Work Orders On Time

Group Size

Average Correlation

-0.20

0.00

0.20

0.40

0.60

0.80

1.00

2

14
Appendix C

Group Expansion Comparison Graphs
Non-TCW Mtnce / Mile (NEG)

Performance

n=1

n=2

Billing Accuracy

Performance

n=1

n=2

Months

1 2 3 4 5 6 7 8 9 10 11 12 13 14

Months

1 2 3 4 5 6 7 8 9 10 11 12 13 14
Billing Timeliness (NEG)

Performance vs. Months

- Months: 1 to 14
- Performance: 0 to 12,000

Graph showing data for two different sets of data points:
- Set 1: n=3, showing an upward trend
- Set 2: n=4, showing fluctuations
- Set 3: n=5, showing fluctuations

Graphs indicate performance over time with different sample sizes.
Labor Exp/Gross Rev (NEG)

Credit Card Net Inc
Billing Accuracy

Performance vs. Months

n=2
n=7

Billing Timeliness (NEG)

Performance vs. Months

n=2
n=7


Huitema, B. E. (March 20, 2001). Personal communication.


