Butler University
Digital Commons @ Butler University

Scholarship and Professional Work - Business | College of Business

7-1-2005

How Groups Produce Higher-Quality Balanced Scorecards than Individuals

S. B. Hughes

Craig B. Caldwell
Butler University, ccaldwel@butler.edu

Kathy A. Paulson Gjerde
Butler University, KPAULSON@BUTLER.EDU

Pam Rouse
prouse@butler.edu

Follow this and additional works at: http://digitalcommons.butler.edu/cob_papers

Part of the Accounting Commons

Recommended Citation
Hughes, S.B., C.B. Caldwell, K. A. Paulson Gjerde, and P. Rouse. "How Groups Produce Higher-Quality Balanced Scorecards than Individuals". Management Accounting Quarterly, volume 6, number 4 (Summer 2005), pp.34-44.

This Article is brought to you for free and open access by the College of Business at Digital Commons @ Butler University. It has been accepted for inclusion in Scholarship and Professional Work - Business by an authorized administrator of Digital Commons @ Butler University. For more information, please contact fgaede@butler.edu.

VOL.5 NO.4
Management Accounting Quarterly
Summer 2005

How Groups Produce **Higher-Quality Balanced Scorecards** Than Individuals

By Susan B. Hughes, Ph.D., CPA; Craig B. Caldwell, Ph.D.; Kathy A. Paulson Gjerde, Ph.D.; and Pamela J. Rouse, CPA

A study shows that groups filter out individuals' poor ideas and incorporate their appropriate ideas. Collectively, groups produce scorecards that contain good quality but primarily mainstream ideas.

EXECUTIVE SUMMARY Many articles explain how to develop a balanced scorecard using groups, but the literature provides little insight about why groups are important. We gathered data from 12 groups involved in developing balanced scorecards to determine how they use information suggested by their members. We found that the groups "filter" individual members' poor ideas and "carry through" their worthy ideas to the group scorecard—although not all poor ideas are filtered and not all good ideas are carried forward. We also found some evidence that groups create innovative ideas but to a lesser extent than filtering and carrying through ideas. Our findings suggest that the outcome of the group process depends on the quality of the potential scorecardobjectives and metrics that group members bring to the discussion. As such, entities that plan to develop a balanced scorecard in a group environment should ensure that the group contains a diverse set of individuals—each with different training, skills, and perspectives—to ensure that the group considers a large pool of good ideas.

Since Robert S. Kaplan and David P. Norton introduced their concept of the balanced scorecard in 1992, the term has become known as a measurement system that links strategic and operating objectives—and lead and lag performance measures—within four areas of an organization: financial, customer, internal business processes, and learning and growth.[1] Scorecard development is generally described as a group process. The balanced scorecard literature, however, provides little insight about why it is beneficial to have groups rather than individuals develop scorecards.

There are two reasons that groups may be preferred. First, groups may produce better scorecards because their team members deliberate and discuss the factors that contribute to the business's success.[2] Second, groups may help scorecard components become accepted and implemented. Yet these two different activities are often combined in books and articles that help businesses plan for and adopt the balanced scorecard.

Research in other management accounting areas often makes a distinction between development and

implementation within an organization. For example, a review of the many research studies that investigated the impact of employee participation on the annual budgeting process suggests that groups add value to the budgeting process only when individuals in the group possess information about the business that is not known by others involved in the budgeting process. This suggests that groups involved in developing a balanced scorecard will add value to its components if they have a deep and clear understanding of what leads to business success. The research conducted on budgeting also found that even when individuals do not make significant contributions to the budget's content, individual participation leads to higher levels of personal satisfaction with the budget and a greater commitment to achieve the budgetary goals.[3] This suggests that an increased level of commitment to the objectives and metrics included in a scorecard may result from employee participation in its development.

Our study separates the development activities from implementation activities and focuses on the impact of groups on a scorecard's development. We based our study on the idea that if groups are commonly recognized as critical to the development and implementation of a scorecard, it is important to understand their advantages and disadvantages. If groups add value to the development process, scorecards developed by groups should be of higher quality than those developed by individuals. If groups add value only to the implementation process, there should be little difference between scorecards developed by individuals and those developed by groups. Within this study we provide an overview of the literature that discusses the potential impact of groups during the scorecard development process and examine the impact of groups on scorecard components.

Group Development of Balanced Scorecards

Developing the balanced scorecard involves identifying organizational metrics that are tied to and can help fulfill a firm's strategy. Advocates of the balanced scorecard believe that business entities rely too much on financial measures and that financial measures by themselves are poor indicators of strategic goal achievement. They argue that financial measures should be "balanced" with nonfinancial considerations. Although traditional financial measures still play a role in evaluating a firm's progress, balanced scorecards give greater consideration to formally tracking and planning for excellent performance in financial as well as nonfinancial areas.

By identifying, monitoring, and learning from a broader range of metrics, firms can benefit from organized learning and measurement in two ways. First, as the process of developing a mission statement provides both process and outcome gains, so, too, can the balanced scorecard process. Second, the tactically oriented metrics developed during the scorecard development process will give firms a way to deconstruct large-scale strategic directions into measurable activities.

Kaplan and Norton first described the scorecard-building process in 1993.[4] They indicated that scorecard construction involves a facilitator and a group of six to 12 executives. The group first reaches consensus on the organization's mission and strategy. Next it defines the appropriate key success factors and identifies the related four or five measures for each scorecard perspective, resulting in a total of 16 to 20 measures. Then, direct subordinates and middle managers are added to the scorecard-building team. These individuals are expected to debate the key success factors and measures. In the final stage, the senior management team finalizes the strategy, key success factors, objectives, and related measures. Kaplan and Norton noted that scorecard building is an interactive process that sometimes takes more than 30 months.[5] Describing the process used by one organization, they list 10 executive team members involved in the initial scorecard development stage. In the next stage, 100 individuals from the top three layers of management discuss the strategy and develop proposed measures used within the various scorecards.

Clearly, many person-hours are devoted to scorecard development during Kaplan and Norton's recommended process. Assuming this time could also be used productively to manage or lead other aspects of the entity's operations, it is important to evaluate if the group time spent developing the scorecard is cost beneficial. Kaplan and Norton suggest that the broad participation of members of the organization in the development of

financial and nonfinancial metrics leads to scorecards that incorporate more information, give employees a better understanding of the firm, and enhance their commitment to the ideas generated. Of these three advantages, only the first is associated directly with the scorecard's development; the other two relate to employees and the scorecard's implementation. Our research focuses on the first advantage and assesses whether groups improve the information content and, therefore, the quality of balanced scorecards.

Team-Based Decision Making

The evidence about team decision making is not uniformly favorable. Some of the problems associated with team decision making include the tendency of groups to favor decision consensus over decision quality, which often is referred to as "group think," and the tendency for groups to make more extreme decisions than individuals, which often is referred to as "group polarization."[6] Despite these recognized problems, group-based decision making continues to enjoy acclaim in both research and practice. As a result, it is important to understand the potential positive effects associated with group decision making.

Similar to the management accounting literature, the literature on groups describes two primary advantages of them. One is the ability to generate higher-quality decisions.[7] The second relates to greater success in implementing decisions.[8] Often these two advantages are blurred, and advantages in implementation become a sufficient reason to use a group-based approach. In our study we detangle the impact of implementation and decision making by focusing exclusively on the quality of decisions generated by individuals and groups.

In order for a group to generate high-quality decisions, three distinct activities must occur. First, marginal suggestions from the group's members must be identified and eliminated. We refer to this process as "filtering." Second, the group members' higher-quality suggestions must be identified and carried through to the group's final output. We refer to this process as "carry through." Third, once high-quality suggestions from the group's members are identified, they must be delineated and refined. We refer to this process as "synergistic idea development." If these three activities occur, then the likely result is a collection of decisions that are better than any single group member could generate by himself/herself.

The ability of a group to generate better outputs than the group's best member is referred to in the literature as an "assembly bonus effect."[9] There is a vigorous debate in the group literature about the ability of a group to perform at a level greater than its most talented member.[10] The variation in research findings appears to be at least partially due to the fact that the presence of an assembly bonus effect is highly sensitive to the type of decision that must be made. Therefore, researchers must test for the existence of an assembly bonus effect in any new or unique group setting. The balanced scorecard is a new and unique setting for this research, and no attempts have been made to establish the existence of an assembly bonus effect in this setting, a fact that further motivated us to conduct this study.

We argue that the groups working to develop a balanced scorecard will be able to generate highly innovative, creative, and high-quality decisions by using the ideas of their individual members. Further, we expect to find an assembly bonus effect by testing for the presence of the three processes described earlier: filtering, carry through, and synergistic idea development.

First we suppose that, during the process of examining individually developed ideas in a group setting, the group will weed out, or filter, those individual decisions that are substandard.[11] The ability of a group to highlight deficiencies in individual thinking is supported by both natural and structured group conflict.[12] In settings involving group conflict, ideas that cannot be supported are revealed and eliminated. The process of discussing various assumptions and ideas forces individuals to reconsider their original thoughts. Thus, we expect that the group decision process will adeptly eliminate substandard items from any list of ideas that the group is considering.

The second supposition we make is that the group will be able to take the input of individual team members and, once the high-quality suggestions are identified, ensure that they are included in the group's final output.[13] As noted earlier, we have described this

process as carry through. Precisely how carry through occurs is not clear. We can, however, observe what happens in the group process and draw conclusions about the extent of carry through likely to occur.

The third supposition we make is that groups will be able to build on and enhance the high-quality decisions that individual group members develop.[14] Because the group members have diverse knowledge, they may become sounding boards for new ideas or suggestions, and the group is likely to see connections that the idea's creator is unable to identify. This ability to improve upon the suggestions of individuals is the synergistic idea development we described previously. The ideas that are generated through this process go beyond the capabilities of any individual. Thus, we expect the assembly bonus effect to help produce high-quality group decisions.

If filtering, carry through, and synergistic idea development occur during the development process, balanced scorecards that groups develop should be of higher quality than those developed by individuals. Further, they should be devoid of low-quality ideas and exhibit some ideas that are more innovative than those that individuals develop. If these results do not occur, the time and energy invested by group members may not provide a corresponding benefit to the scorecard's content.

METHODOLOGY

To test our suppositions, we had 46 MBA students participate in a balanced scorecard development exercise. The exercise was a graded part of the course assignments for the MBA capstone class, a course generally taken at the end of the MBA program. The rationale for using these students was twofold. By using students, we were able to assign an identical task to multiple groups for the purposes of comparison. In a corporate setting, task assignment of this sort probably would be prohibitively expensive and impractical. Moreover, because these students had significant work experience, they had the necessary skills, training, and knowledge to develop a credible balanced scorecard.

We gave all participants a packet of information about the same U.S. public corporation. We chose this entity because of its relatively simple structure, clear mission, and readily identifiable product. The packet included corporate strategic objectives, product information, details about a recent acquisition, and information about the company's work environment and commitment to its communities.

Each student received a blank scorecard with the information packet. The scorecard had four rows, one for each area of the balanced scorecard, and three columns. The columns provided boxes for objectives, lead measures, and lag measures for each scorecard area. Lead measures identify metrics that are useful in predicting that the objective will be achieved; lag measures are used to determine if the objective was met. This design resulted in 12 cells within the scorecard.

We used two phases so we could assess the extent to which groups filter, carry through, and develop ideas. First we asked each participant to individually develop a balanced scorecard for the company. The participants were asked to limit their suggested measures to no more than 24, which is similar to the number of measures Kaplan and Norton recommend. In the next phase, the individual participants were placed in three- to four-person groups and asked to complete a balanced scorecard as a team, using their individual scorecards as the basis for discussion.

In order to establish a benchmark scorecard for the exercise, a panel of four faculty members from the areas of accounting, leadership, and economics individually developed a balanced scorecard for the company using the same packet of information the students used. These faculty members then worked as a team to complete a balanced scorecard. This scorecard, shown in Table 1, became the basis for evaluating the students' scorecards.

The 46 individuals proposed 1,258 ideas; the 12 groups proposed 376 ideas. We used content analysis to convert the narrative ideas into a form suitable for data analysis. Content analysis relies upon the ability of at least two independent coders to read and assign the narrative—in this case, each idea—to a specific category. The coders used the following three categories to classify the individual and group ideas:

1. **An inappropriate idea** is one that does not match or approximate the faculty panel's ideas and is considered to be inappropriate for this scorecard cell.

Table 1: Model Scorecard

Objectives	Lead Measures (To gauge if objective will be met)	Lag Measures (To determine if objective was met)
FINANCIAL		
Strong, balanced growth	Number of new products developed Increase market share Appropriate acquisitions	Double sales More than double profits Percent of sales by product category Integrate gains/savings
Fair return to shareholders		Return on equity Earnings per share Increase dividend/share Stock price exceeds $30 per share during fiscal year
Maintain independence		Maintain/reduce debt Reduce capital expenditures to below 4.5% of sales Maintain liquidity ratios at existing levels Focus on cash flow opportunities from operations rather than from other areas
CUSTOMER		
Grow market share of existing brands	Number of new retail accounts added Number of new industrial accounts added	Number of retail and industrial accounts retained Increase sales (domestic and foreign) Increase profits
Introduce new products	Number of new products under development	
Maintain high-quality, "market leader" brand image	Number of brands that are marketplace leaders within their respective category Number of new "icon brands" acquired Fair pricing (price relative to competitors' prices)	
Increase customer awareness	Ongoing support of sponsors and strategic partners Advertising spending	Unsolicited customer comments (by volume and quality)
INTERNAL BUSINESS PROCESSES		
Improve product quality	Defect rate or measure of conformity	Sales growth
Encourage innovation in terms of product and process	Number of new products under development Number of continuous improvement projects Increase production efficiencies (cost minimization or fixed at current percent of cost of goods sold)	Maintain capital budget at 4.7% of sales Increase margins Establish inventory turnover measures
Effectively integrate new acquisitions	Number of poorly performing accounts eliminated Number of poorly performing products eliminated Number of new products under development Number of new logical accounts acquired	
LEARNING AND GROWTH		
Continue to be values-driven company	Awareness of core values Number of appropriate hires Training and development cost Level or rate of job growth	*Fortune* 500 ranking Employee turnover rate Absenteeism rate Number of internal hires

2. **An appropriate idea** is one that matches or approximates the faculty panel's ideas of this scorecard cell.

3. **An innovative idea** is one that does not match or approximate the faculty panel's ideas for this scorecard cell but represents an appropriate, interesting, or thoughtful idea.

Accordingly, two individuals working independently coded each idea contained in the individual and group scorecards as inappropriate, appropriate, or innovative. The codings were compared, and the coders agreed with each other 92.3% of the time. Disagreements between the coders were resolved by a third coder. According to the definitions and standards developed in this process, the first two coders reviewed the data a third time to ensure consistency across scorecard cells.

To make a further distinction about those items coded as inappropriate, we separated them into two distinct sets of ideas. If the idea was deemed inappropriate as a result of content error (e.g., the participant identified an objective or measure that was not applicable to the organization or simply did not make sense), it was coded as a content error; if the idea was deemed inappropriate because it placed an idea within an inappropriate cell, it was classified as a placement error. As before, two people coded these 993 ideas. Their coding agreed for 85% of the ideas, and the third coder resolved the differences for the other 15% of the ideas.

Findings

We began evaluating the impact of group participation on the objectives, lead measures, and lag measures of the four balanced scorecard sections by developing summary statistics of the inputs and outputs of the scorecard development process. These results are summarized in Table 2. One finding that is immediately apparent is that the number of ideas developed at the individual level and used as the basis for group discussion varied widely. For example, at least one group began the group discussion with only three lead or lag measures for various cells of the scorecard. Another group began its discussion of financial lag measures with 17 individual ideas. The group results were more clustered, ranging from one idea per cell to five ideas per cell. We also found that the number of ideas by scorecard section (financial, customer, internal business process, and learning and growth) was approximately equal as was the number of items suggested for objectives, lead measures, and lag measures.

Table 2: Balanced Scorecard Ideas Developed at Group and Individual Levels

Groups evaluate the ideas individuals bring to the group discussion and reduce them in all scorecard areas and topics.

	Number of Ideas Individuals Developed			Number of Ideas Groups Discussed		
	Objectives	Lead Measures	Lag Measures	Objectives	Lead Measures	Lag Measures
Financial						
Average	10.17	7.75	9.67	2.00	2.42	2.83
Range	5–13	3–15	5–17	1–3	2–4	1–4
Customers						
Average	9.67	9.08	7.08	2.50	2.75	2.67
Range	6–15	3–16	4–12	2–3	1–4	1–4
Internal Business Processes						
Average	9.67	8.00	9.00	2.67	3.08	2.50
Range	5–14	3–15	5–14	1–4	2–5	1–4
Learning and Growth						
Average	9.17	8.00	7.58	2.67	2.50	2.50
Range	6–14	3–15	5–13	1–5	1–4	1–4

Table 3: Change in Number of Appropriate Ideas

When comparing the individuals' suggestions to the groups' scorecards, the percentage of inappropriate ideas decreases and the percentage of appropriate ideas increases.

	PERCENTAGE REDUCTION OF INAPPROPRIATE IDEAS IN EACH SCORECARD SECTION (IN PERCENT)			PERCENTAGE INCREASE IN APPROPRIATE IDEAS IN EACH SCORECARD SECTION (IN PERCENT)		
	Objectives	Lead Measures	Lag Measures	Objectives	Lead Measures	Lag Measures
Financial	20	20	24	10	18	21
Customers	22	30	3	14	26	5
Internal Business Processes	16	7	22	16	3	24
Learning and Growth	25	10	14	23	11	12

Next we summarized the coded individual and group ideas to evaluate whether groups improved the quality or enhanced the creativity of the scorecard ideas. One difficulty in comparing the individual and group results is the difference in the number of ideas included. This occurred because the groups were evaluating the ideas of three or four individuals to develop one proposed scorecard and were asked to develop scorecards in which the number of measures was limited to 24. To facilitate comparison between the individual and group ideas, we converted the number of ideas to percentages.

We found that 777 of the 1,258 individual responses (62%) were coded as inappropriate either for that area of the scorecard or for the scorecard in general. We identified only 40 ideas (3%) that were coded as innovative, and the majority of these were included within the lead measures. The remaining 441 ideas (35%) were appropriate for their areas of the scorecard. We found that appropriate and innovative ideas outnumbered inappropriate ideas only in the areas of internal business processes objectives and financial and customer lag measures. These results suggest that individuals relied upon common business objectives and metrics—those that may be applicable to many different businesses—rather than developing innovative and thoughtful ideas appropriate or specific to their company.

Within the group scorecards, there were 191 (51%) appropriate and 22 (6%) innovative ideas from among 376 group ideas. Both of these percentages were higher than those found in the individual scorecards. One-half of the innovative ideas occurred within the lead measures. Only within three scorecard cells—financial objectives, financial lead measures, and learning and growth lead measures—were the majority of the group ideas coded as inappropriate. Within the internal business process lag measures, the ideas were evenly divided between those coded as inappropriate and appropriate.

We found that the percentage of inappropriate ideas declined from the individual to the group results within all 12 scorecard cells. We also found that the percentage of ideas coded as appropriate increased from the individual to group levels within all 12 cells. The details of these results are shown in Table 3. The percentage of group ideas coded as innovative was greater than that at the individual level in most cells, although the number of innovative ideas continued to be less than 10% of all group ideas. These results indicate that the group process appears to have improved the overall quality of the ideas, filtering the inappropriate ideas and improving the concentration of innovative ideas. The group process also added a limited number of unique and high-quality ideas within the scorecard sections.

Group Selection of Individual Ideas

To further explore how the groups selected ideas, we analyzed the individual and group ideas in each group to identify those individual ideas in each cell that were also among the group's ideas in the same cell. This analysis allowed us to track the frequency with which the various groups included ideas suggested by an individual member from his or her scorecard. We were particularly interested in determining if groups had the ability to distinguish between inappropriate and appropriate ideas included in the individual scorecards, which would be indicated by the elimination of inappropriate ideas and the incorporation of appropriate ideas within the group scorecards.

The groups appear to have engaged in a fairly effective sorting process. The number of appropriate individual ideas that carried forward to the group level exceeded the number of inappropriate ideas in all 12 of the scorecard cells, as shown in Table 4. Few innovative ideas were carried forward, so we focus our discussion on the inappropriate and appropriate responses.

The results also indicate that groups differed in the extent to which they relied on individual ideas carried forward in determining the group ideas included within the 12 cells. The majority of group ideas within the financial, internal business process, and learning and growth objectives came from individual ideas, as summarized in Table 5. The lead measures for three scorecard areas, however, were largely developed within the groups. Within the lag measures, the majority of ideas within two cells were the result of individual ideas and within two cells were the result of ideas developed by the groups. The results reveal that although the groups relied on the individual ideas when developing the financial objectives and lag measures (adding only three of 25 and six of 34 ideas, respectively), in all other areas they tended to supplement the individual ideas in varying and greater amounts.

Table 4: "Carry Through" of Individual Ideas to the Group

Groups carry through a lower percentage of inappropriate individual ideas and a higher percentage of appropriate individual ideas in all scorecard areas.

	INDIVIDUAL IDEAS CARRIED THROUGH (IN PERCENT)		
	Objectives	**Lead Measures**	**Lag Measures**
FINANCIAL			
Inappropriate Ideas	16	13	9
Appropriate Ideas	48	22	34
CUSTOMERS			
Inappropriate Ideas	7	8	8
Appropriate Ideas	34	25	22
INTERNAL BUSINESS PROCESSES			
Inappropriate Ideas	16	6	6
Appropriate Ideas	30	37	36
LEARNING AND GROWTH			
Inappropriate Ideas	10	12	12
Appropriate Ideas	38	37	55

Table 5: Ideas Carried Through by Cell

Group scorecard cells include different levels of individually developed ideas.

	INDIVIDUAL IDEAS AS A PERCENT OF GROUP IDEAS ADOPTED IN THE GROUP SCORECARD		
	Objectives	**Lead Measures**	**Lag Measures**
FINANCIAL	84	41	82
CUSTOMERS	50	35	38
INTERNAL BUSINESS PROCESSES	78	43	50
LEARNING AND GROWTH	66	70	63

Inappropriate Ideas

As explained in the methodology section, we further analyzed the inappropriate ideas to determine if they reflected content error or classification differences. As noted, a content error occurs when an individual or group identified an objective or measure that was not applicable to the organization or simply did not make sense. A classification error occurs when a potentially appropriate idea appeared in an incorrect cell. If the groups added value to the development of the scorecard items, we would expect that more of the inappropriate items result from classification differences and fewer result from content errors. Again, because of the larger number of ideas included in the individual ideas than in the group ideas, we used percentages to determine if the quality of the scorecard items improved from individuals to groups.

Our results indicated that the groups had the ability to filter individuals' content errors from the group's financial objectives and that the groups eliminated many content errors from the objectives of the other three scorecard sections. The groups, however, exhibited varying success in eliminating individual ideas classified as content errors from the lead and lag measures of the group scorecards. These results are summarized in Table 6. They clearly indicate that groups do not effec-

Table 6: Eliminating "Content" Errors*

Groups do not effectively reduce individuals' inappropriate content.

	CONTENT ERRORS (IN PERCENT)		
	Individuals	Groups	Did Groups Improve the Content?
OBJECTIVES			
Financial	7.6	0.0	Yes
Customers	20.0	23.1	No
Internal Business Processes	30.9	11.1	Yes
Learning and Growth	14.5	8.3	Yes
LEAD MEASURES			
Financial	35.7	37.5	No
Customers	31.0	50.0	No
Internal Business Processes	31.5	33.3	No
Learning and Growth	23.1	27.8	No
LAG MEASURES			
Financial	46.7	40.0	Yes
Customers	35.1	38.5	No
Internal Business Processes	25.6	12.5	Yes
Learning and Growth	25.0	25.0	No Change

* Content errors occur when an individual or group identifies an objective or measure that is not applicable to the organization or does not make sense.

tively filter inappropriate ideas identified by individual group members from all areas of the group scorecards.

Group Improvement of Ideas

The analyses discussed in the previous paragraphs suggest that the group process adds value to the balanced scorecard development process by reducing the percentage of inappropriate ideas included within the cells and by including more of the appropriate individual ideas within the group scorecard. We find, however, that the groups continue to carry many inappropriate ideas within their proposed scorecards. The analyses presented so far are based upon totals and percentages determined within each of the scorecard cells.

To further investigate the influence of individual ideas and input to the group scorecard ideas, we analyzed the results of each group. Because we found few individual and group ideas coded as innovative, we combined the two categories of good ideas (appropriate and innovative) into one "good" idea total per group per cell. By combining these two categories, we could compare the number of inappropriate and good ideas individuals developed in each group and compare those numbers with the inappropriate and good ideas included in the group's output within each scorecard cell. Consistent with the prior analyses, we relied on percentages for comparative purposes. We tabulated the percentage of good to total answers among the individual ideas of each group and the corresponding percentage of good to total group answers within the four scorecard areas. The difference between the individual and group percentages of good answers formed a "percentage improvement." We ranked each group based on the percentage of good answers included in the group scorecards.

The results indicate that the quality of the individual scorecards was relatively low as determined by the percentage of good ideas to total ideas at the individual level. Only one or two groups have more than 50% of their total ideas come from individual ideas within each section of their scorecards. When all ideas from all groups are combined, the individual ideas are less than 50% of the total ideas.

The group results reveal that most groups developed scorecards in which the percentage of good ideas exceed 50%. More than 50% of the ideas within each scorecard section of at least seven groups were good. More than 50% of the total ideas of eight groups were good. These results, compared with group results in the aggregate, indicate that the improvement within the balanced scorecards was not generated by a few groups but was widespread across many groups.

We also calculated the percentage improvement in good responses from individual to group responses for each group in each of the four scorecard areas and the combined improvement over all four sections. This calculation revealed that the quality improved from the individual to group scorecards within all four scorecard sections within almost all of the 12 groups. The results indicated that the quality declined for only one group in the financial, customer, and learning and growth sections and for only two groups within internal business processes. When the results of the four areas were combined, all group scorecards improved over those of their individual members.

Relative Performance

To determine the relative performance of the groups, we ranked them by the percentage of good ideas determined at the group level from high to low, using a one-to-12 scale, adjusted for ties. The rankings were computed within each scorecard section and for the combination of all four of the scorecard's sections. Within the combined scorecards, the percentage of good ideas ranged from a high of 72.4% to a low of 34.4%. The six top-performing groups increased the percentage of good ideas from 18.3% to 34.4% over the performance of their individual members.

We also summed the group rankings by scorecard section and found the two top-performing teams in the combined-section approach were similarly ranked under the individual-section approach. In fact, five of the top six teams in the combined results were in the top six of the individual-section approach. As such, the combined results appeared to provide a reasonable approximation of the performance levels of the groups. A review of the details underlying the analysis did not provide any evidence that a specific group improved its performance by incorporating ideas coded as innovative. The results indicated that innovative ideas were widely dispersed across the groups and that only one group submitted a scorecard that included only ideas coded as inappropriate and appropriate.

Good-Quality, Mainstream Ideas

Our initial questions were:

- Are poor ideas generated by individuals eliminated in the group decision-making process (filtering)?
- Are high-quality ideas generated by individuals identified and embraced as a worthy idea for the team's final report (carry through)?
- Are the ideas generated by the team likely to include highly innovative ideas (synergistic idea development)?

The answer to all three questions is a qualified yes. Of these three effects, filtering appears to be the most prevalent, suggesting that the primary benefit of group decision making in a balanced scorecard context is eliminating inappropriate ideas, not introducing appropriate or innovative ideas. Although these results may be partially attributed to the fact that the participants in our study met only once as a group, thereby limiting their opportunity to brainstorm and debate, the results still suggest the potential limitations of group decision making in a balanced scorecard setting.

In essence, the team dynamic resulted in good-quality but fairly mainstream ideas. This is a critical finding given that the purpose of strategic thinking is to generate solutions that provide competitive advantage. Firms are said to have a competitive advantage when they have implemented a strategy that is unique, novel, or difficult to replicate.[15] If the team process results in scorecard objectives and metrics that are mainstream, they are likely to be similar to those of others in the same industry. In this case, rather than contribute to an entity's competitive advantage, the scorecard may simply replicate objectives and measures across different companies, thereby reducing its positive benefits.

From an organizational perspective, our findings suggest that the information or ideas fed into the process at the beginning stage are critical to the development process. In other words, while the group decision-making process may effectively filter out inappropriate ideas, it does not seem to add a significant number of interesting, insightful ideas. Thus, our results highlight the importance of group composition. In particular, groups will not produce a sufficiently large pool of good ideas without a diverse set of individuals, each with different training, skills, and perspective. ■

Susan B. Hughes, Ph.D., CPA, is professor of accounting at Butler University in Indianapolis. She can be reached at (317) 940-9843 or shughes@butler.edu.

Craig B. Caldwell, Ph.D., is visiting professor of management at Butler University. He can be reached at (317) 940-8154 or ccaldwel@butler.edu.

Kathy A. Paulson Gjerde, Ph.D., is associate professor of economics at Butler University. She can be reached at (317) 940-8413 or kpaulson@butler.edu.

Pamela J. Rouse, CPA, is lecturer in accounting at Butler University. She can be reached at (317) 940-8816 or prouse@butler.edu.

ENDNOTES

1 Robert S. Kaplan and David P. Norton, "The Balanced Scorecard—Measures that Drive Performance," *Harvard Business Review*, January/February 1992, pp. 71-79.

2 The literature on teams and groups establishes clear differences between the two constructs. Teams are a variation of groups, but groups are not always teams. For the purposes of this article, however, we use the terms "team" and "group" interchangeably.

3 For an extensive review of the literature on participative budgeting, see Mark A. Covaleski, John H. Evans, III, Joan L. Luft, and Michael D. Shields, "Budgeting Research: Three Theoretical Perspectives and Criteria for Selective Integration," *Journal of Management Accounting Research*, 2003, pp. 127-141.

4 Robert S. Kaplan and David P. Norton, "Putting the Balanced Scorecard to Work," *Harvard Business Review*, September/October 1993, pp. 134-147.

5 Robert S. Kaplan and David P. Norton, "Using the Balanced Scorecard as a Strategic Management System," *Harvard Business Review*, January/February 1996, pp. 75-85. Other evidence of the use of groups in scorecard development is found in Andra Gumbus and Bridget Lyons, "The Balanced Scorecard at Philips Electronics," *Strategic Finance*, November 2002, pp. 45-49.

6 For "group think," see Brian Mullen, Tara Anthony, Eduardo Salas, and James E. Diskell, "Group Cohesiveness and Quality of Decision Making: An Integration of Tests of the Groupthink Hypothesis," *Small Group Research*, May 1994, pp. 189-204; for "group polarization," see Craig McGarty, John C. Turner, Michael A. Hogg, Barbara David, and Margaret S. Wetherell, "Group Polarization as Conformity to the Prototypical Group Member," *British Journal of Social Psychology*, March 1992, pp. 1-20.

7 This view is widely held. For example, see the work of Steven P. Robbins, *Organizational Behavior*, 10th ed., Prentice-Hall, Upper Saddle River, N.J., 2003.

8 For an early version of this argument, refer to Norman R. Maier, "Assets and Liabilities in Group Problem Solving," *Psychological Review*, October 1967, pp. 239-249.

9 Although considerable work has subsequently been done on the concept, the original idea of the assembly bonus effect can be traced to Barry E. Collins and Howard Guetzkow, *A Social Psychology of Group Processes for Decision Making*, John Wiley & Sons, New York, N.Y., 1964.

10 For the two perspectives on this issue, see the work of R. Scott Tindale and James R. Larson, "It's Not How You Frame the Question, It's How You Interpret the Results," *Journal of Applied Psychology*, February 1992, pp. 109-110; and Larry K. Michelsen, Warren E. Watson, Albert Schwartzkopf, and Robert H. Black, "Group Decision Making: How You Frame the Question Determines What You Find," *Journal of Applied Psychology*, February 1992, pp. 106-108.

11 Carsten K. W. De Dreu and Michael A. West, "Minority Dissent and Team Innovation: The Importance of Participation in Decision Making," *Journal of Applied Psychology*, December 2001, pp. 1191-1201.

12 David M. Schweiger, William R. Sandberg, and James W. Ragan, "Group Approaches for Improving Strategic Decision Making: A Comparative Analysis of Dialectical Inquiry, Devil's Advocacy, and Consensus," *Academy of Management Journal*, March 1986, pp. 51-71.

13 *Ibid*.

14 Wesley M. Cohen and Daniel A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly*, March 1990, pp. 128-152.

15 Michael E. Porter, *Competitive Advantage: Creating and Sustaining Superior Performance*, Free Press, New York, N.Y., 1985.