Children’s Learning Groups: A Study of Emergent Leadership, Dominance, and Group Effectiveness
RYOKO YAMAGUCHI 671

Problem-Solving Team Behaviors: Development and Validation of BOS and a Hierarchical Factor Structure
SIMON TAGGAR and TRAVOR C. BROWN 698

The Provision of Effort in Self-Designing Work Groups: The Case of Collaborative Research
NATHAN BENNETT and ROLAND E. KIDWELL, Jr. 727

Reexamining Teamwork KSAs and Team Performance
DIANE L. MILLER 745

Index 767
POLICY: SMALL GROUP RESEARCH is an international and interdisciplinary journal presenting research, theoretical advancements, and empirically supported applications with respect to all types of small groups. Through advancing the systematic study of small groups, this journal seeks to increase communication among all who are professionally interested in group phenomena.

MANUSCRIPTS should be submitted (4 copies) to Charles Garvin, School of Social Work, University of Michigan, Ann Arbor, Michigan 48109. Manuscripts should be typed in accordance with the guidelines set forth in the 1994 edition of the Publication Manual of the American Psychological Association. A copy of the final revised manuscript saved on an IBM-compatible disk should be included with the final revised hard copy. Submission of a manuscript implies commitment to publish in the journal. Authors submitting manuscripts to the journal should not simultaneously submit them to another journal, nor should manuscripts have been published elsewhere in substantially similar form or with substantially similar content. Authors in doubt about what constitutes prior publication should consult the editor.

SMALL GROUP RESEARCH (ISSN 1046-4964) is published six times annually—in February, April, June, August, October, and December—by Sage Publications, 2455 Teller Road, Thousand Oaks, CA 91320; telephone (800) 818-SAGE (7243) and (805) 499-9774; fax/order line (805) 375-1700; e-mail order@sagepub.com; http://www.sagepub.com. Copyright © 2001 by Sage Publications. All rights reserved. No portion of the contents may be reproduced in any form without written permission of the publisher.

Subscriptions: Annual subscription rates for institutions and individuals are based on the current frequency. Prices quoted are in U.S. dollars and are subject to change without notice. Canadian subscribers add 7% GST (and HST as appropriate). Outside U.S. subscription rates include shipping via air-speeded delivery. Institutions: $415 (within the U.S.) / $439 (outside the U.S.) / single issue: $80 (worldwide). Individuals: $88 (within the U.S.) / $112 (outside the U.S.) / single issue: $24 (worldwide). Orders from the U.K., Europe, the Middle East, and Africa should be sent to the London address (below). Orders from India and South Asia should be sent to the New Delhi address (below). Noninstitutional orders must be paid by personal check, VISA, or MasterCard.

Periodicals postage paid at Thousand Oaks, California, and at additional mailing offices.

This journal is abstracted or indexed in ASSIA (Applied Social Science Index & Abstracts), Automatic Subject Citation Alert, Current Contents, Educational Administration Abstracts, Health Instrument File, Human Resources Abstracts, Linguistics & Language Behavior Abstracts, Psychological Abstracts, PsycINFO. Research Alert, Sage Family Studies Abstracts, Sage Urban Studies Abstracts, Social SciSearch, Social Sciences Citation Index, and Social Work Abstracts and is available on microfilm from University Microfilms, Ann Arbor, Michigan.

Back Issues: Information about availability and prices of back issues may be obtained from the publisher’s order department (address below). Single-issue orders for 5 or more copies will receive a special adoption discount. Contact the order department for details. Write to the London office for sterling prices.

Inquiries: All subscription inquiries, orders, and renewals with ship-to addresses in North America, South America, Australia, China, Indonesia, Japan, Korea, New Zealand, and the Philippines must be addressed to Sage Publications, 2455 Teller Road, Thousand Oaks, CA 91320, U.S.A.; telephone (800) 818-SAGE (7243) and (805) 499-9774; fax (805) 375-1700; e-mail order@sagepub.com; http://www.sagepub.com. All subscription inquiries, orders, and renewals with ship-to addresses in the U.K., Europe, the Middle East, and Africa must be addressed to Sage Publications Ltd., 6 Bonhill Street, London EC2A 4PU, England; telephone +44 (0)20 7374 0645; fax +44 (0)20 7374 8741. All subscription inquiries, orders, and renewals with ship-to addresses in India and South Asia must be addressed to Sage Publications Private Ltd., P.O. Box 4215, New Delhi 110 048, India; telephone (91-11) 641-9884; fax (91-11) 647-2426. Address all permissions requests to the Thousand Oaks office.

Authorization to photocopy items for internal or personal use, or the internal or personal use of specific clients, is granted by Sage Publications, for libraries and other users registered with the Copyright Clearance Center (CCC) Transactional Reporting Service, provided that the base fee of $0.50 per copy, plus 10¢ per copy page, is paid directly to CCC, 21 Congress St., Salem, MA 01970. 1046-4964/2001 $0.50 + .10.

Advertising: Current rates and specifications may be obtained by writing to the Advertising Manager at the Thousand Oaks office (address above).

Claims: Claims for undelivered copies must be made no later than six months following month of publication. The publisher will supply missing copies when losses have been sustained in transit and when the reserve stock will permit.

Change of Address: Six weeks’ advance notice must be given when notifying of change of address. Please send the old address label along with the new address to ensure proper identification. Please specify name of journal.

POSTMASTER: Send address changes to Small Group Research, c/o 2455 Teller Road, Thousand Oaks, CA 91320.
This study explores the importance of the group context in the emergence of leadership, dominance, and group effectiveness in children’s cooperative learning groups. Using achievement goal orientation as a framework, six groups performed a math task under a mastery condition, whereas four groups performed a math task under a performance condition. Under the performance condition, group members exhibited more dominance and negative behaviors, whereas under the mastery condition, group members exhibited more leadership and positive behaviors. Also, under the performance condition, groups were not as effective in cooperatively completing the math task because of negative communication, member dissonance, and isolation, whereas under the mastery condition, groups were more effective, demonstrating positive communication, group cohesion, and a shared responsibility in completing the math task. Implications for classroom practice are discussed.

Leadership is one of the most popular topics of group research, attributing leadership to group effectiveness (Borg, 1957; Chemers, 2000; Cohen, Chang, & Ledford, 1997; Hare & O’Neill, 2000), group cohesion (Dobbins & Zaccaro, 1986; Hurst, Stein, Korchin, & Soskin, 1978; Neubert, 1999; Schriesheim, 1980), and ultimately, group and organizational success (Bass, 1985; Collins &
TABLE 1: Group Composition and Overview of Results in Children’s Learning Groups Study

<table>
<thead>
<tr>
<th>Group ID (school level)</th>
<th>Group Membersa (student ethnicity)</th>
<th>Emergent Leadership or Dominance</th>
<th>Group Effectiveness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mastery group 1 (elementary)</td>
<td>Aaron (African American) Eric (Caucasian) Natalie (Caucasian)</td>
<td>shared leadership</td>
<td>did not finish task (time crunch)</td>
</tr>
<tr>
<td>Mastery group 2 (middle)</td>
<td>Sue (Asian American) Bill (Caucasian) Larry (Caucasian)</td>
<td>shared leadership</td>
<td>finished task/ correct</td>
</tr>
<tr>
<td>Mastery group 3 (elementary)</td>
<td>Jim (Caucasian) Jenny (Caucasian) Erin (Caucasian)</td>
<td>shared leadership</td>
<td>finished task/ correct</td>
</tr>
<tr>
<td>Mastery group 4 (middle)</td>
<td>John (Caucasian) Catie (Caucasian) Sonia (Caucasian)</td>
<td>shared leadership</td>
<td>finished task/ correct</td>
</tr>
<tr>
<td>Mastery group 5 (elementary)</td>
<td>Aine (Asian American) Becky (Caucasian) Efrem (African American)</td>
<td>shared leadership</td>
<td>finished task/ correct</td>
</tr>
<tr>
<td>Mastery group 6 (elementary)</td>
<td>Brandon (Caucasian) Zack (Caucasian) Nora (Caucasian)</td>
<td>shared leadership</td>
<td>did not finish task (lost materials)</td>
</tr>
<tr>
<td>Performance group 7 (elementary)</td>
<td>Rob (Caucasian) Harvey (African American) Tina (Caucasian)</td>
<td>dominance: Rob</td>
<td>did not finish task</td>
</tr>
<tr>
<td>Performance group 8 (middle)</td>
<td>Gina (Caucasian) Liz (Caucasian) Cody (Caucasian)</td>
<td>dominance: Gina</td>
<td>finished task/ correct [Gina completed task]</td>
</tr>
<tr>
<td>Performance group 9 (elementary)</td>
<td>Heather (Caucasian) Michelle (Caucasian) Adam (Caucasian)</td>
<td>dominance: Heather</td>
<td>did not finish task</td>
</tr>
<tr>
<td>Performance group 10 (elementary)</td>
<td>Tomika (African American) Mary (Caucasian) Chris (Caucasian)</td>
<td>dominance: Chris</td>
<td>did not finish task</td>
</tr>
</tbody>
</table>

*a. All names are pseudonyms.

Porras, 1994; Kotter, 1988). However, leadership studies, including informal and emergent leadership (Sorrentino, 1973; Sorrentino & Field, 1986; Wheelan & Johnston, 1996; Wheelan & Kaeuser, 1997), have focused primarily on adults. Research in children’s
leadership, particularly emergent leadership, is limited (Edwards, 1994). However, anecdotal evidence of children’s emergent leadership is apparent in everyday life. For example, in learning groups, certain students under certain conditions emerge as the group leader. In sports groups, regardless of skill, certain children emerge as leaders of their team. In playgroups, children emerge as leaders, directing play themes and games.

The emergence of leadership is an important, yet missing, factor in how children learn in groups. Previous research on leaderless groups has found that leaders emerge within the group through certain achievement-related behaviors and motives (Bass, 1949; French & Stright, 1991; French, Waas, Stright, & Baker, 1986; Sorrentino, 1973; Sorrentino & Field, 1986; Stein, 1977; Stein, Geis, & Damarin, 1973). Behaviors, such as engagement, facilitating, soliciting opinion, organizing, and record keeping, distinguished emergent leaders from other group members (French & Stright, 1991). Hence, students who emerge as leaders are those who take an active role in the group and learning process. Emergent leadership, therefore, is a critical factor in understanding group effectiveness and achievement.

However, whereas emergent leadership is associated with prosocial group behaviors, scholars have researched dominance in children's group behavior (Parten, 1933; Pigors, 1933, 1935; Savin-Williams, 1979, 1980; Savin-Williams, Small, & Zeldin, 1981). Pigors (1933) defines leadership as the guidance of others toward a common goal, whereas domination is the forcing of others, by assertion of superiority, to perform acts that further the dominator’s private interests. Also known as the “bully,” dominators may actually decrease interaction and learning in group collaboration (LaFreniere & Charlesworth, 1983; Segal, Peck, Vega-Lahr, & Field, 1987; Trawick-Smith, 1988).

Hence, how can educators influence the emergence of leadership, not dominance, in cooperative learning groups? Educators can influence the context in which the students learn. More specifically, research on achievement goal orientation provides evidence for two types of group contexts: mastery and performance goal orientations. Although students are required to achieve in school,
mastery and performance goal orientations differ on the purpose of achieving. Mastery goal orientation refers to a focus on learning and improving, whereas performance orientation refers to a focus on competition and social comparison (Ames & Archer, 1988; Maehr & Midgley, 1996). Research has shown that performance goals tend to be associated with maladaptive patterns of cognition, affect, and behavior (Middleton & Midgley, 1997; Midgley, Anderman, & Hicks, 1995; Midgley, Arunkumar, & Urdan, 1996; Midgley, Feldlaufer, & Eccles, 1988), and mastery goals are associated with adaptive patterns (Maehr & Midgley, 1996; Maehr, Midgley, & Urdan, 1992; Pintrich & de Groot, 1990; Pintrich & Schrauben, 1992). In school settings, mastery goal orientations focus on the task per se: progress in learning and mastering a skill, intrigue with an unanswered question. In school settings with a performance goal orientation, the focus is often on performing competitively and demonstrating who is “smarter.” Instead of influencing the motivational group context through external rewards (Slavin, 1996), goal orientation influences the group context through emphasizing the purpose and beliefs about learning (Ames & Archer, 1988; Maehr & Midgley, 1996; Pintrich & Schunk, 1996).

The purpose of this study is to explore emergent leadership, dominance, and group effectiveness under different learning conditions. Using achievement goal orientation theory as a framework, the research questions are as follows:

1. Does the learning condition of the group influence the emergence of leadership or dominance?
2. Does the learning condition of the group influence group effectiveness?

METHOD

PARTICIPANTS

The 30 participants in this study (53% female; 23% minority) were drawn from a larger sample of 133 fourth-, fifth-, and sixth-grade students across five elementary school classes and four mid-
middle school classes in a metropolitan area in the midwest. Students from each class were grouped into three-person learning groups. Ten triads were randomly selected from the larger sample to facilitate videotaping. All names have been changed to pseudonyms to ensure confidentiality; however, gender and ethnicity of the students were preserved.

**DATA COLLECTION AND ANALYSIS**

The data used for this analysis were taken from a larger study on motivation in cooperative learning groups, during the 1997-1998 school year (Linnenbrink, Hruda, Haydel, Star, & Maehr, 1999). The study was conducted in collaboration with the school improvement team, led by the principal of each school.

The students took part in a cooperative math activity. The math activity follows McGrath’s (1984) intellective task, where the goal is to determine the correct answer. An intellective task relies on the cooperation or interdependence of group members to find the correct answer (McGrath, 1984; Straus, 1999). However, although students can benefit from cooperating and sharing information, any one of the group members can dominate and complete the math task individually. Students worked together for 30 minutes to plan a hypothetical field trip to Chicago, staying within a budget of $2,400 while fulfilling the requirements or “rules” of planning. Specifically, with the $2,400, the students had to plan and budget for travel, hotel, lunch, dinner, and two fun activities per day. Each student in the group was given tickets, or information, about travel, hotel, lunch, dinner, and fun activities. The main math questions involved figuring out how many days they can spend in Chicago without going over budget, how much the total trip cost, and how much money they had left over. The intellective task, therefore, not only involved determining the correct answers, but reaching consensus on the answers and the planning. Scratch paper and calculators were provided to all students. Each group completed one answer sheet, the “Trip Planner,” on which the effort of the group as a whole can be evaluated.
If students completed the Chicago trip before the 30-minute period, they were asked to complete another similar math activity, which involved planning a trip to Cancun. While continuing to budget for travel, hotel, lunch, dinner, and two fun activities per day, the Cancun math activity was more difficult because it involved converting dollars into pesos. The second math activity was created for two reasons. First, the Cancun activity allowed for differences in ability and age group. Second, the Cancun activity allowed for classroom management, where it ensured that all groups worked throughout the 30 minutes. However, the main focus of the math activity, and the main measure of group effectiveness, was the completion of the Chicago trip.

For the math activity, all students within their classroom were assigned to three-person learning groups. All groups of students within a given classroom were given either mastery or performance instructions. These instructions followed patterns commonly employed in experimental studies of this nature (Butler & Neuman, 1995; Graham & Golan, 1991). For the mastery goal orientation, the class was instructed to complete the math task to the best of their knowledge, but the purpose of the math task was learning and improving. Students in the mastery condition received the following instructions:

We made this math activity to see if it would help you learn and understand how to use math when you do things outside of school. This is not a test; we’re simply interested in making a math activity that will help kids learn. This should be interesting and fun to do.

Throughout the cooperative math activity, students were reminded that the focus was on learning, understanding, and improving.

For the performance goal orientation, the class was instructed to complete the math task to the best of their knowledge, but the purpose of the math task was to test their math ability and see who was “best” at math. Students in the performance condition received the following instructions:

This math activity is designed as a way of testing how well (elementary or middle school) students can use math outside of school.
When you finish the activity, we’ll be looking at your papers to see how much you know about using math outside of school. We’ll also be sharing this information about how well you did with your teacher.

Throughout the cooperative math activity, students were reminded that the focus was on doing better than other groups and to see “who was the best in math.”

Two groups from each classroom were randomly selected to videotape their group interaction. However, eight videotapes were deleted from the analysis because of technical problems with the audio-visual recording. Thus, the data used for this study focused on the videotaped interaction of 10 three-person learning groups; 4 groups were performance oriented, 6 groups were mastery oriented. Field observations were conducted prior to, during, and following the cooperative math task, and always in the student’s own classroom. The videotapes were fully transcribed.

Qualitative methodology, specifically grounded theory, was chosen for this study to understand social interaction patterns related to children’s learning groups. By being field focused, detail oriented, and giving voice to students, qualitative methodology allows for an in-depth understanding of the subtle processes of emergent leadership and dominance in children’s collaboration learning groups (Bogdan & Biklen, 1998; Creswell, 1998; Eisner, 1991; Eisner & Peshkin, 1990; Strauss & Corbin, 1990).

The qualitative software, NVIVO, was used to manage and code the transcripts. In coding and analyzing the transcripts, Glaser and Strauss’s (1967) constant comparative method, also referred to as open coding (Strauss & Corbin, 1990), was used. Open coding is the process of breaking down, examining, comparing, conceptualizing, and categorizing data (Strauss & Corbin, 1990). In the first phase of coding, mastery and performance group contexts were eliminated from each transcript to minimize bias in coding and analyzing the data. After the group contexts were eliminated, the transcripts were coded in random order. Each transcript was coded, or “chunked” (Miles & Huberman, 1984, 1994), into units of physical activity, such as fighting, giving orders, calculating, and organizing. These units of physical activity included sequences of verbal
and nonverbal interactions. Each chunk consisted of one to six speaking turns per student, and there were 6 to 16 chunks (average 8.5 chunks) of coded group interaction per transcript. The chunks were not multiply coded. After analyzing the units of physical activity, the domains were further refined into the following categories: group effectiveness, dominance, and leadership. The second phase of coding involved reintroducing the mastery and performance group contexts and analyzing the patterns that emerged. The third phase of coding involved expanding the three categories into subcategories to examine the different kinds or nuances of group effectiveness, dominance, and leadership under mastery and performance group contexts. Overall, coding the transcripts was a reiterative process of reducing and expanding categories to create a coherent story of children’s learning groups under mastery and performance group contexts.

RESULTS

1. Does the learning condition of the group influence the emergence of leadership or dominance?

In analyzing the group interactions during the cooperative math activity, the mastery or performance learning condition greatly impacted the emergence of leadership and dominance. In performance groups, one member dominated his or her way into the top position by bullying and taking over the group process. Although previous studies on adults have found gender differences in emergent leadership, where women emerge as social leaders and men emerge as task leaders (Eagly & Karau, 1991; Kolb, 1999), there were no gender differences in emergent leadership or dominance for the fourth-, fifth-, and sixth-grade students in this study.

For Performance Group 7, Rob pushed, yelled, and blamed his way to take over the group process. In one instance, Rob was ordering Harvey to add a set of numbers, which Harvey was reluctant to do. When Tina offered to add the numbers instead, Rob became upset and insisted that they all add together. In another example of
Rob’s dominance, he physically pushed Harvey away, yelling, “Get out of here!” when Harvey was trying to verify Rob’s calculations. Instead of viewing the math task as a cooperative and democratic process, Rob behaved as if his group members worked for him, as evidenced when he would order his group members to raise their hand when he or she had a question about the math task. In addition, Rob consistently antagonized Tina by excluding her from the math process or calling her names. Toward the end of the math task, Tina realized that the group did not fill in the answer sheet. Because Rob was doing all the math calculations, Tina asked him to fill out the answer sheet. Initially, Rob ordered Tina to fill out the answer sheet. When she refused, he retorted, “Look. It says, ‘Group members.’ You Retard! Everybody’s got to fill it out.”

In Performance Group 8, Gina made all the decisions, as evidenced by the following example:

Liz: Do you guys want to go to the Shedd Aquarium?
Gina: How much does it cost?
Liz: $54.
Gina: Ok.
Cody: Get the cheapest!

Cody continued to suggest selecting the cheapest items, often disagreeing with Gina. Although Cody and Gina would argue, Gina continued to dominate the group process and the math task because Liz always sided with Gina. Cody, therefore, was often isolated and silenced.

In Performance Group 9, Heather dominated the verbal conversation, to the dismay and irritation of her group members. She often snapped at her group members (sometimes literally) and eventually horded all the materials, including all three calculators. The following is an example of Heather’s hording:

Heather: For lunch, for lunch, what do you want for lunch? Let’s go to Hard Rock.
Michelle: Here’s the answer sheet.
Heather: We don’t use that yet! Do you have any hotel tickets? Give me one of those!
Toward the end of the group process, the other group members were clearly annoyed and irritated with Heather's dominance, especially when Heather accused them of “not helping.”

Performance Group 10 was an unfortunate example of “the blind leading the blind.” It was clear that Chris did not understand the math task. However, he dominated the group by overpowering, shouting, and hording the materials. Tomika, who understood the directions and the concepts involved in the math task, tried to emerge as the leader throughout the group interaction. Unfortunately, as this example shows, Chris shouted and over-voiced his way to the top position.

Tomika: We don’t need Day 2. See, it says, “If you need it.”
Chris: Hold up! (And starts calculating) $1752. What was the hotel?
Mary: (Looks for ticket) $420.
Chris: Oh well, we’re still on our budget any way, where’s Day 2?
Tomika: (Gives Chris a dirty look) We don’t need Day 2!
Chris: Yea . . . We do!
Tomika: It says, “If needed!” And we don’t need it!
Chris: We plan two days. Now we’re figuring this out (And gets worksheet).

The group interaction of Performance Group 10 was dominated by such verbal tug-of-war. However, toward the end of the math activity, the verbal tug-of-war became a physical one, as the example below illustrates.

Chris: (In a panic) Where’s the sheet!?! We got to fill it up!
Mary: (Calmly) It’s in the envelope.
Chris: NO! We got to fill it out!
Tomika: (Takes it out and starts writing on the answer sheet) How much is Day 2?
[Chris tries to take the answer sheet away from Tomika. They are in a tug-of-war with the answer sheet. Finally, Tomika backs down and loosens her grip on the answer sheet. Chris quickly snatches it away.]
Chris: No, I got to see it! (He starts filling in the answers while the girls quietly watch him).
To Tomika’s credit, she continued to try to contribute and lead in the cooperative math task. However, Chris never backed down, verbally or physically, and continued to dominate the group process.

In mastery groups, leadership emerged in all students, where all members shared the responsibility of completing the task. More specifically, each member led the group at different times. Leadership in the mastery group not only focused on the math task, but it also focused on group cohesion and building positive group interactions.

In Mastery Group 1, Natalie often emerged as the leader. However, toward the end of the math task, Aaron directed the group processes. Eric, who tended to be off-task, was also able to lead some parts of the math activity. The following example shows how leadership shifted from Eric, then to Aaron, who suggested starting over, and finally to Natalie.

**Eric:** The first day we have to have one of those or we’re not going anywhere!
**Aaron:** What’s that?
**Natalie:** Oh, yea! (Agrees with Eric) We have to have that the first day ‘cause . . . Plus that, Aaron! (Smiling and laughing with Aaron)
**Eric:** Yea, the first day we’re not going anywhere. We’re just gonna go around on our own buying stuff if you don’t put that in there! (Clearly excited that he knows how to do the math activity)
**Aaron:** We gotta start over!
**Natalie:** Oh yea. Aaron, no Aaron, no. (Pause) Ok, fine.
**Eric:** We’re going to the museum!
**Natalie:** Ok, we’re starting all over. (Agrees with Aaron)

Although the verbal interaction between Eric, Aaron, and Natalie seemed negative, the nonverbal interaction and behaviors indicated excitement in the activity, playfulness, and group camaraderie.

Mastery Group 2 was another team where leadership not only emerged in all members but also was shared and distributed throughout their interaction. In almost a round-robin approach, Larry indicated to the group that he would add a set of numbers and wanted Bill to organize the tickets. Bill agreed and, while he organized the tickets, asked Sue to subtract a set of numbers. Sue in turn
agreed and gave her answer to the group and suggested what to do next. This round-robin approach of leadership was evidenced throughout their interaction.

In Mastery Group 3, Jim, Jenny, and Erin equally focused on the math task and group cohesion. For example, Jim suggested, “Erin, you pick your (tickets). Jenny, you pick yours. And I’ll pick mine.” When two group members disagreed, the third smoothed things out, as evidenced by Erin:

   Jenny: No, it’s add.
   Jim: No, it’s take away, duh.
   Erin: Jim, don’t say “duh,” it’s not nice.

Jim immediately apologized and reiterated his motivation to work together to complete the math task.

In Mastery Group 4, John made it a point not to overtly take over the group process. Rather, he encouraged Catie and Sonia to take part in the group decisions. For example, John suggested that each member be in charge of one aspect of the math task.

   John: Each one of us has to cover something—so Sonia, you be like food and fun things, and I’ll be like arriving and leaving. Ok? You know, that’s just my suggestion . . . Ok, you wanna do that? (To Sonia) Two and two? (Looks at Catie and asks) What do you want to do?
   Catie: I don’t know? I guess travel and hotel.
   John: Travel and hotel? (Looks at Sonia) What do you want?
   Sonia: I’ll trade you, arriving for fun things. (To Catie) So what do you want?
   John: (To Catie) And I got fun things and leaving.

Although Catie was initially reluctant to participate, John’s encouragement and insistence on group decisions made Catie an equal participant toward the end of the math activity.

Mastery Group 5, although especially focused on finishing the math task, also worked on improving group cohesion, as evidenced in this example:

   Becky: If they take away . . .
Aine: What’s Hard Rock Cafe? (To Efrem)
Efrem: Shhh, listen to Becky (To Aine).
Becky: If we take away the 2 nights at the hotel, we have this much left (Shows calculator to Efrem).

For Mastery Group 6, because Zack understood how to do the math activity, he often directed his group members with regard to the math activity. However, Mastery Group 6 shared other responsibilities, such as asking the researcher questions, calculating the math activity, and writing down the answers, as evidenced by the following example:

Nora: $858 (Adding tickets for the first day).
Brandon: This one’s $1056! (Adding tickets for the third day).
Zack: Somebody add up Day 2!
Brandon: I will. Where’s Day 2?
Nora: I got it right here.

Overall, dominators emerged in performance groups, where an individual overtook and overpowered the math and group processes. However, in the mastery groups, group members took turns taking the leadership role. In addition, whereas one member emerged as math leaders, focusing on the calculations, other members emerged as social leaders, focusing on group cohesion. These emergent leadership roles were not static in the mastery groups. Both math and social leadership roles were shared among members in the mastery groups.

2. Does the learning condition of the group influence group effectiveness?

The mastery or performance learning condition impacted group effectiveness. In performance groups, group members exhibited more negative group interactions and communication, such as social loafing, fighting, and blaming each other during the group process. Such interactions inhibited performance groups to cooperatively complete the math activity.

The communication became stagnated in performance groups. In Performance Group 7, Rob, Harvey, and Tina struggled to com-
communicate with each other. Rob and Harvey were so competitive with one another that they often ended up in power struggles. For example, when Harvey gave an answer, it was met with disdain from Rob.

Harvey: That’s $46.
Rob: How do YOU know!??
Harvey: 'Cause I used a calculator.
Rob: Well, then figure it out! 30 plus . . .
Harvey: See, that’s $46!
Rob: 46! We got lunch ticket, afternoon fun thing, we got this, and we got this.

Whenever Rob and Harvey got into a power struggle, Harvey often asked a researcher to help resolve their differences. In all, Harvey asked a researcher for help 12 times during the group interaction. Unfortunately, Rob, Tina, and Harvey did not seek help from each other.

In addition, Rob often silenced Tina. At the end of the group interaction, Tina had her hands resting on her face, just watching Rob and Harvey continue their power struggles. At one point in the interaction, Rob accused Tina of “stealing the tickets,” called her names, and suggested that she was not as “smart” as the other members, as evidenced by this example:

Harvey: Second day, I get to plan.
Rob: No, I get to plan!
[Tina grabs the envelopes for Second Day and smiles at Harvey and Rob.]
Rob: He gets to plan, he’s the second smartest at this table!
Tina: No he’s not! (But clearly looking demoralized).

It is not surprising that Performance Group 7 did not finish the math task. Harvey and Rob were so busy trying to prove to each other “who was smarter” that they neglected to complete the math task.

Performance Group 8 was able to complete the math task. However, completing the math task was not a cooperative group effort as intended. Gina finished most of the math task, not the group. She decided the plans of the trip, wrote down the answers on scratch
paper, and asked the researcher questions about the task. Her dominance in the group interaction and the math activity led to social loafing, inactivity, and boredom in Liz and Cody.

In Performance Group 9, poor communication was exhibited when Heather completely ignored her fellow group members. This group split into two factions. Because Heather ignored her group members, Adam and Michelle ended up trying to do the math activity themselves. Although the math activity was designed to be cooperative, Heather took over the math task and tried to complete it herself, as evidenced by the following:

Michelle: (To Heather) Where are we going to eat?  
Heather: Don’t know. I’m almost done. Ok! I only have $500 to go! I made it! I have to add this up now!  
Michelle: What are you doing Heather?  
Heather: I have to get this out before we do anything! [Heather continues to calculate feverishly, ignoring Adam and Michelle.]  
Michelle: (To Adam) What is she doing? (Michelle and Adam both shrug).  
Researcher: Ok, times up. Please put everything in the envelopes and I’ll come to pick them up.  
Michelle: We don’t know the answer or nothing!  
Heather: Well, I got MY answers. Here, just write your names!

Although Heather completed one part of the math task, Performance Group 9 was unable to finish the math task. Although Michelle and Adam did their best to take part in the math task, they were clearly bored during the math activity due to Heather’s dominance. When the 30-minute period was over and the students were handing in their math activity, Heather whole-heartedly complained to her group and the researcher.

Heather: We didn’t get half an hour! This ain’t right! Guess we can’t go to Cancun no more! I don’t think we should have put our names on it! Just kidding, we should have (half-smiling at the researcher). This ain’t fair! I can’t even do that that fast! At least some people got something, WE didn’t! (Looking around the classroom).

Michelle and Adam did not engage Heather in her tirade. Michelle and Adam did not seem to care that their group did not finish the
math task. Rather, they simply looked relieved that the math activity was over.

Like the other performance groups, Performance Group 10 was unable to effectively communicate with one another to complete the math task, as evidenced below.

Tomika: No, you guys, this is our ticket home.
Chris: It's $430, now we need the hotel!
Tomika: Time out! Everybody adds this now.
Chris: (Pulls ticket out of Tomika's hand) Let me read one of those!
Mary: (To Tomika) That's $360.
Chris: Wait a minute! How about if we try one night and maybe that'll cut the price (and starts to calculate).
Tomika: Wait! Chris! (Looks at Mary).
Mary: (To Tomika) Just let him do it. Just until he adds it up and we'll just do the rest.

While Mary and Tomika tried to work on the math task on their own, Chris continued to hoard the materials and give orders.

Chris: Ok, now give me everything else. The hotel and the bus . . .
[Tomika and Mary try to ignore him and work on their own.]
Chris: (Irate that he is being ignored) Keep the hotel! [Mary continues to talk to Tomika] We've got things taken care of . . . the hotel and the bus.
Tomika: (Confronting Chris) Where's the bus?
Chris: I don't know.
Tomika: There is no bus, there's the plane.

Because Chris continued to dominate his way through the math task, Mary and Tomika had built much animosity toward Chris. With a few minutes remaining in the math task, Performance Group 10 tried to complete the answer sheet to no avail.

Chris: How much do we have left over? I think it was $648.
Mary: Yea.
Tomika: No, we added more (Snatches his calculator and adds).
Chris: Let me see this! (And snatches calculator back).
Mary: Let me see this! (And takes the worksheet).
Chris: (After a while of calculating) I need the sheet back.
[Tomika starts her own calculations while Mary watches Chris.]
Chris: How much was . . . [He tries to take the sheet that Tomika is working on. Mary comes to Tomika’s defense by giving Chris a very dirty look.]

Unfortunately, Performance Group 10 was so busy snatching calculators, tickets, worksheets, and answer sheets from one another that they never finished the math task in all the chaos.

In mastery groups, group members exhibited more positive group interactions and communication, hence, completing the math task effectively and cooperatively. Students in the mastery groups not only completed the math task on time but also enjoyed the cooperative learning experience. Similar to performance groups, students in the mastery group faced disagreements. However, even when there were disagreements or a group member was off-task, students in the mastery group relied on one another for help. Suggestions were met with enthusiasm from other group members. The following example from Mastery Group 1 shows the dynamic flow of the conversation:

Eric: Day 1 we got the museum, the trip, and the food. That’s all we need, ok.
Natalie and Aaron (in unison): Ok.
Natalie: I need the trip tickets!
Eric: Got it.
Natalie: I need something to do after you.
Aaron: Day 2.
Natalie: We need something in the morning and afternoon.

Mastery Group 1, however, did not finish the math task. First, Eric was often off-task during the math activity. Second, the group agreed to completely redo the math task twice. While such willingness to redo the math task was admirable, they simply ran out of time.

Mastery Group 2 exhibited effective communication, even when there was disagreement on the task. Larry, Bill, and Sue were able to diffuse any possible power struggles and relied on the instructions for clarification. This group was self-reliant, only asking questions to the researcher when they could not find the answer in the instructions, as this example illustrates:
Larry: We can leave in the morning. Then, we don’t have to spend for lunch.
Bill: We can’t leave in the morning.
Larry: Yea, we can.
Bill: Ok.
[Sue looks in the directions and finds that they do leave in the morning.]
Sue: See in the rules. (Reads instructions aloud) You leave early in the morning.

However, Larry, Bill, and Sue were confused if leaving in the early morning constituted paying for the night. When they were all stumped, they asked the researcher for clarification. Because of their self-reliance and effective communication, Mastery Group 2 correctly finished their math task.

Mastery Group 3 correctly finished their math task through effective communication and shared responsibility. When something was amiss, instead of fighting or blaming, the group members helped one another, as this example shows:

Jenny: We have $146 left.
Jim: Huh? We did this wrong.
[All group members look at the math calculations. Erin takes over and redoes Jim’s calculations. Jim and Jenny look on.]
Erin: We need one more night at the hotel and eat at the cheapest place.

Mastery Group 4 positively communicated with one another by alleviating trepidation in the task and relying on one another throughout the math activity, as evidenced by the following example. Mastery Group 4 correctly completed the math task.

John: This is going to be hard, we’re all going to have to work together.
Sonia: It’s easy; we’ve got a calculator.
[John and Sonia pick up the calculators.]
Catie: (Reads instructions aloud) But, I don’t know (Looking worried).
John: Everybody give their ticket to the ticket person. All right. Hotel person, fun things . . . (Organizes tickets).

Mastery Group 5 experienced frustrations, anxieties, and disagreements, especially when time was running out to complete the math task. However, Mastery Group 5 correctly finished the math
task because they had an incredible ability to positively communicate with each other and even joke about their disagreements. Efrem was excited about the math task from beginning to end, often stating that the math task was “business stuff.” This excitement also motivated his group members, as evidenced by the following example:

Efrem: Ok. We have $990 left ’cause I added it up. We only need things to do there. We’ve got breakfast covered.
Becky: So we’ve got lunch and fun things.
Efrem: Fun things.
Aine: I’ll do lunch, ok?
Efrem: This is all business stuff, right here! (Clearly excited about the activity)
Becky: What would they most like to do in the afternoon?
Efrem: They can go to the beach.
Aine: Yea! The beach!

Mastery Group 6 had a difficult time getting started because they had trouble understanding the math task. After receiving instructions again from the researcher, they were able to do the math task on their own. Instead of constantly asking the researchers for help, they were able to help and rely on each other. The following example illustrates their efficient and effective communication:

Zack: We have to add this all up . . . to see how long we can stay.
Nora (Picks up the calculator): 2 take away 18, plus 30, plus 420, plus 60, plus 330 . . . what do I do over here?
Zack: (Goes over to her end of the table) Day 2, here. Which place do you want to go to eat?
Brandon: $850.
Zack: Ok, put this in the envelope.
Brandon: What? All this?
Zack: Yea.
Nora: Ok.

Mastery Group 6, however, did not complete the math task because Nora misplaced the tickets. With the help of two other researchers, the group ended up looking for their tickets for the last 10 minutes of the allotted time.
In mastery groups, group members exhibited more math strategies than performance groups. Math strategies included viewing the math task as “fun,” trying out different math combinations, and selecting the cheapest tickets to solve the math task. In Performance Group 8, Liz indicated that the math task was “fun.”

Liz: Do you guys want to do it (math task) again for fun?
Gina: No! We have another one to do anyway!
Cody: [Agrees with Gina, and ignores Liz’s comment.]

Unfortunately for Liz, her comment was immediately dismissed and she remained quiet throughout the rest of the group interaction. However, groups in the mastery condition were much more receptive to viewing the math task as fun. For example, in Mastery Group 2, all the group members validated Sue’s comment that the math activity was fun.

Sue: This should be fun!
Bill: Yea, this should be fun!
Larry: Yea, 500 pesos is probably like, $200 here!

In Mastery Group 3, Jim was obviously excited about the activity, which also motivated his group members, Jenny and Erin. Even before the researcher finished the instructions, Jim loudly whispered to his group, “This’ll be fun!” Jim asked the researcher for a weather report of the area so his group can plan the field trips accordingly. Jim continued to whisper to his group, “You got the tickets and the envelopes. Ok you guys! I have the afternoon, you’ve got the morning, you’ve got fun things.” While Jenny and Erin agreed with Jim’s suggestions and were also excited about the math task, they reminded him to finish listening to the researcher, who was still giving instructions to the class.

In the mastery condition, students were much more willing to be adventurous and try out different math combinations to answer the math problems. For example, in Mastery Group 5, Efrem concluded, “So we can only stay 2 nights.” Before Efrem can write the answer on the “Trip Planner,” Becky interrupted and proclaimed, “No, hold on! I’m going to see if we can stay 3 nights!” Such will-
The mastery groups were able to figure out that in order to stay as long as possible in the field trips, the students had to find the “cheapest” and most affordable itinerary. Mastery students were suggesting ways to “save money.” In Mastery Group 1, Eric’s strategies to save money were on the creative side, suggesting that to save money for more fun activities, they can “go on a diet to eat less food,” and “sleep in a rental car and eat pizza.” In Mastery Group 5, Efrem’s suggestion to save money was to “bring a bag lunch,” instead of eating out.

Students in the mastery group figured out that to stay as many days as possible, they needed to save money by selecting the cheapest items for their field trip. In Mastery Group 2, Bill, Sue, and Larry were focused on saving money throughout the math activity. For example, when Larry wanted to go to an expensive restaurant for lunch, Sue stated, “We should save money for the fun activity.” They all agreed to a cheaper lunch.

Mastery Group 3 was focused on planning the field trips as cheaply as possible, but they also checked each other’s math calculations. Throughout their group interaction, the focus was on find-
ing the cheapest items, without regard to their preference in fun activities or restaurant. This strategy was drastically different from the performance groups, where they focused more on preference of fun activities than solving the math problem.

In performance groups, students did not readily understand the concept of saving money, or budgeting, to plan the trip to Chicago. Instead of selecting the cheapest items, or saving money for the fun activities, performance group students got caught up in organizing the tickets, trading the tickets with each other, and debating what they wanted to do in Chicago. Whereas the tickets offered such fun activities as “go to the beach” and “go to the zoo,” students in the performance group were often off-task, talking about their family vacations in Chicago.

Overall, performance groups were unable to complete the math task cooperatively. One performance group, however, completed the math task because of the dominator’s efforts. Negative communication and ineffective math strategies in the performance groups inhibited the cooperative completion of the math task. Mastery groups were able to complete the math task cooperatively. Two mastery groups were unable to complete the math task, although the circumstances for incompletion were not the same as performance groups. One mastery group lost their materials, and the other mastery group kept redoing the math task. Positive communication and effective math strategies in the mastery groups facilitated the cooperative completion of the math task.

CONCLUSION

This study shows that the learning condition plays an important role in the emergence of leadership, dominance, and group effectiveness. Mastery groups exhibited more prosocial leadership, focusing on the math task and group cohesion. The emergence of leadership in the mastery groups was distributed and shared among group members. In addition, mastery groups were more effective in solving the math task, as evidenced by their math strategies and positive communication. Performance groups exhibited more
dominance, where one student, regardless of race or gender, emerged to overpower the group process. Performance groups were less effective in solving the math task because friction within the groups created ineffective communication and math strategies. Students in the mastery groups enjoyed the math activity and asked the researchers to come to their classroom again. Unfortunately, students in the performance groups, particularly those who were dominated, looked frustrated, stressed, and bored during the math activity. Finally, it is important to note that mastery and performance groups faced similar problems, such as confusion on the math task and occasional friction between group members. The differences between the groups were their strategies and interactions to solve these problems. By emphasizing learning and improving, mastery groups were able to focus on the task at hand. By emphasizing competition and social comparisons, performance groups experienced friction and fighting within their groups and were not able to concentrate on completing the math task.

These results have direct implications for practice. Advocates of cooperative learning have touted the values of group learning experiences (Slavin, Madden, Dolan, & Wasik, 1995), without taking into consideration the learning environment. This study shows that the group condition is an important factor in successful cooperation and learning. In addition, this study shows that emergent leadership is a positive feature of group interaction and learning, whereas domination is a negative component of group interaction and learning.

Teachers and educators can influence the emergence of leadership, and limit dominance, in group learning situations. Specifically, teachers can prime the students into a mastery orientation by emphasizing learning and improving as goals. Unfortunately, with an increased emphasis on testing and standards, schools often emphasize performance goals, focusing on social comparison, student ability, and competition. However, research has shown that mastery goal orientations provide better learning environments for students, particularly disadvantaged students (Baden & Maehr, 1986; Maehr & Midgley, 1999; Midgley et al., 1996; Midgley & Edelin, 1998). In addition, the results of this study show that mas-
tery goals are beneficial to promote learning and equity in cooperative groups. To conclude, children’s learning groups provide opportunities for students to learn from each other (Slavin, 1996), as well as learn social interaction skills. However, for learning groups to be an effective teaching technique, teachers must also shape the group environment to focus on mastery goal orientations.

REFERENCES


Ryoko Yamaguchi is a Ph.D. candidate in educational administration and policy at University of Michigan, Ann Arbor. Her research focuses on collaboration and group dynamics, leadership, motivation, and diversity.
Development and Validation of BOS and a Hierarchical Factor Structure

SIMON TAGGAR
York University

TRAVOR C. BROWN
Memorial University of Newfoundland

This is among the first studies to develop a typology of performance-relevant team member behaviors from actual observations of intact teams ($N = 94$) completing a variety of problem-solving tasks. Through a critical incident analysis, behavioral observation scales (BOS) were developed using confirmatory factor analysis. The correlation between BOS and team performance was significant. The typology supports and adds to previous typologies of teamwork behavior. Subsequent cross-validation of the BOS involving 176 team members provided further support for the BOS developed. Next, we used our data to psychometrically evaluate Stevens and Campion’s hierarchical factor structure. Hierarchical confirmatory factor analysis provided support for Stevens and Campion’s model.

Over the past decade, ample evidence has been collected to demonstrate that European and North American employees often do not work in isolation from each other but work in teams (Guzzo & Shea, 1992; Latham & Seijts, 1997; Lawler, 1986; Wimmer, McDonald, & Sorensen, 1992). The misalignment of performance appraisal research, which continues to focus on individual jobs that are isolated from one another, and current organizational contexts such as team-based organizations have been noted by several

AUTHORS’ NOTE: The authors would like to thank Gary P. Latham for his helpful comments on an earlier draft on this article as well as the feedback received from Richard Kettner-Polley and two anonymous reviewers.
researchers (Latham, Skarlicki, Irvine, & Siegel, 1993; Lawler, 1986). As teams are becoming more common in today’s organizations, it is critical that researchers examine performance appraisal procedures that are useful for evaluating individual performance in team settings (Murphy & Cleveland, 1991; Saavedra & Kwun, 1993; Stevens & Campion, 1994). As such, three purposes of the present study were to (a) develop a comprehensive performance appraisal instrument for assessing individual team member behavior on problem-solving tasks, (b) assess the validity of this instrument by correlating individual team member behavior with team performance on team problem-solving tasks conducted over a 13-week period, and (c) cross-validate the instrument on a second sample.

Some, although not many, researchers have developed categorical schemes to delineate the diverse types of behaviors found in different teams (e.g., Cannon-Bowers, Tannenbaum, Salas, & Volpe, 1995; Hyatt & Ruddy, 1997; Stevens & Campion, 1994). Teamwork typologies are differing beliefs about what team members do, or should do, for the team to perform effectively. They are shaped by the team task, team composition, norms, and political priorities. Therefore, teamwork typologies can be considered a product of organizational and situational characteristics (i.e., reward structure, environmental uncertainty, and resources), task characteristics (complexity, organization, and type), as well as work characteristics (work structure, team norms, communication flow) (Cannon-Bowers et al., 1995). As typologies develop, one may expect some diversity and commonality among the categories contained in different typologies. These commonalities and differences can be attributed to (a) the purpose of the team, as some categories may not be needed for all team tasks; (b) scholars using differing principles of classification; (c) certain scholars having their own rhetorical purposes; and (d) categories being conceptualized at differing levels of abstraction—that is, they differ on bandwidth and fidelity. As a result, Cannon-Bowers et al. (1995) argue that,

the literature is plagued with inconsistencies, both in labels and in definitions of teamwork . . . researchers often use different labels to
Thus, team researchers should now examine (a) what constructs are conceptualized in a particular typological hierarchy, (b) how these constructs are related to each other and other typologies, and (c) why these relationships exist. The concept of teamwork will become more meaningful and useful if researchers can find a method to examine relationships among these typologies and identify the relative importance of various teamwork constructs. Otherwise, we risk a literature that lacks consensus and development as researchers each pursue their own work without any attempt to build on and integrate the work of others.

Therefore, our fourth purpose was to integrate our work with that of other researchers in the field. In particular, we investigate the higher order factor structure of problem-solving teamwork constructs through hierarchical confirmatory factor analysis.

This article is organized into three parts. First, we review existing typologies of teamwork behaviors. Second, we describe the development of a 46-item teamwork inventory. Finally, we present and discuss results of hierarchical confirmatory factor analysis of data from 480 team members.

TYPOLOGIES OF TEAMWORK BEHAVIOR

To date, few studies have attempted to identify a comprehensive set of performance-relevant behaviors that define effective teamwork. Stevens and Campion (1994), in the earliest and most cited of these studies, offered a synthesis of the group- and organizational-level literature. They inferred 14 generic individual-level knowledge, skills, and abilities (KSAs). These KSAs fell into two categories: interpersonal and self-management. Interpersonal KSAs were further divided into the subcategories of conflict resolution, collaborative problem solving, and communication categories. Self-management KSAs were grouped into two subcategories: goal-setting and performance management, and planning and task coordina-
tion. Subsequent studies that assessed the validity of these KSAs revealed mixed results.

Stevens and Campion (1999) developed a pen and pencil test based on their KSAs. They found, using samples of wood workers \((N = 70)\) and box plant workers \((N = 72)\), that the test had criterion validity with supervisor and peer ratings. In contrast, Miller (1999), in a field study involving 176 students, failed to find a statistically significant correlation between self-assessments of team members’ performance, using Stevens and Campion’s (1994) KSAs, and the team’s performance on team problem-solving tasks. These findings suggest that Stevens and Campion’s KSAs may not generalize to all team tasks, in particular, those concerning group problem solving. Moreover, although teamwork behavior is conceptualized as the behaviors necessary to be an effective team member (Stevens & Campion, 1999), there may not be a statistically significant relationship between individual demonstration of these KSAs and overall team effectiveness or performance.

Hyatt and Ruddy (1997) conducted roundtable sessions with work group members and managers to develop a rating scale to measure work group effectiveness. They found that effective work groups were high in process focus, work group support, goal orientation, work group confidence, customer orientation, and interpersonal work group processes. However, the authors concluded that the factor structure contained a high degree of redundancy and suggested that additional research be conducted. Moreover, they suggested that their findings require replication partly due to limitations inherent in conducting research in a field setting where group performance data cannot be collected in a controlled and consistent manner.

Cannon-Bowers et al. (1995) compiled an analysis of the behaviors necessary for effective team performance based on an extensive literature review. In all, Cannon-Bowers et al. identified and defined eight skill dimensions (Performance monitoring and feedback, Interpersonal relations, Leadership, Coordination, Adaptability, Decision making, Communication, Shared situational awareness) and 25 subskills. A possible limitation of this typology is that it was constructed at the team level. Stevens and Campion
(1994, 1999) have stressed the importance of individual teamwork behaviors given that many human resource systems are based on the individual rather than on the team.

A potential weakness of these preceding typologies is that they were not developed through a systematic job analysis. For the past four decades, researchers have advocated the use of behavioral measures of performance developed through a job analysis that systematically identifies the KSAs necessary for successful performance (Campbell, Dunnette, Lawler, & Weick, 1970; Latham et al., 1993; Smith & Kendall, 1963). A behavioral measure that has been studied extensively in the literature is Latham and Wexley’s (1977) behavioral observation scales (BOS).

The advantages of BOS are at least three-fold. First, previous research has shown that BOS have test-retest reliability, inter-observer reliability (Latham & Wexley, 1977; Latham, Wexley, & Rand, 1975; Ronan & Latham, 1974), and construct validity, as demonstrated in a double-cross validation study of loggers where outcome criteria (i.e., productivity) was positively and significantly related to BOS scores (Latham & Wexley, 1977). Second, BOS improve performance when used in conjunction with goal setting (Dossett, Latham, & Mitchell, 1979; Latham, Mitchell, & Dossett, 1978). Potential explanations for this goal setting effect were found by Tziner and Kopelman (1988) who, in a field study of Israel Airport employees, found that employees appraised using BOS had significantly higher levels of goal clarity, goal acceptance, and goal commitment than did those appraised using a graphic rating scale. Third, BOS are composed of behavioral referents that are under the control of the ratee and are observable. As such, BOS focus the rater’s attention on pertinent behaviors and, therefore, conform to Wherry and Bartlett’s (1982) recommendations concerning methods of minimizing bias in performance ratings.

Recently, other studies have developed BOS that included teamwork behaviors (e.g., Brown & Latham, 1999; Sue-Chan & Latham, 1999). However, as training interventions rather than the development of a teamwork BOS were the focus of these studies, they were not comprehensive typologies in that they omitted sev-
eral of Stevens and Campion’s (1994) KSAs. Further, the primary focus of these studies was at only one level of analysis—the individual. As a result, these studies did not examine the relationship between individual demonstration of BOS behaviors and overall team performance.

Given the limitations of the previous studies, we sought to develop comprehensive BOS from actual observations of intra-group team member behavior over a variety of problem-solving tasks and then validate these BOS by correlating individual team member behavior (i.e., total score on the BOS) with team performance on these group problem-solving tasks. The goal of this part of the study was to (a) empirically investigate the appropriateness of Stevens and Campion’s teamwork KSAs when the team task is primarily problem solving, (b) add behavioral specificity to existing typologies, including that of Stevens and Campion, when the team task is primarily problem solving, (c) add new dimensions to these typologies that are relevant to problem-solving tasks, and (d) demonstrate a statistically significant relationship between individual teamwork behavior and team performance, hence, addressing a concern raised by Stevens and Campion (1994) and Miller (1999).

HIERARCHY OF TEAMWORK BEHAVIOR

Our first goal was to assess the appropriateness of Stevens and Campion’s KSAs. Specifically, we used hierarchical confirmatory factor analysis (HCFA) to test a hierarchy of their teamwork construct. What distinguishes HCFA from other forms of factor analysis is that researchers must specify a model with first-order and higher order factors before analyzing the data. The modeling of constructs as second-order and third-order factors permits broad, more interesting constructs to be included in the model and yet accommodates the rigorous evaluation of unidimensionality that follows from the specification of a multiple-indicator measurement model. Jöreskog (1970) introduced confirmatory second-order factor analysis, following Thurstone’s (1947) treatment of second-order exploratory factors. A detailed description of HCFA is
beyond the scope of this article but is available elsewhere (e.g., Gerbing, Hamilton, & Freeman, 1994; Marsh, 1987; Mulaik & Quartetti, 1997; Rindskopf & Rose, 1988). Instead, the following section presents our conceptualization of Stevens and Campion’s (1994) KSA hierarchical factor structure as part of a hierarchical factor model useful for organizing factors generated from a problem-solving team.

Figure 1 depicts the model hypothesized for this study. This model consisted of 2 third-order factors, 5 second-order factors, and 46 first-order factors that were derived from a job analysis. The 2 third-order factors and 5 second-order factors represent Stevens and Campion’s (1994) 2 categories (interpersonal, self-management) and 5 subcategories (conflict resolution, collaborative problem solving, communication, goal-setting and performance management, planning and task coordination), respectively. The first-order factors applicable to problem-solving teams were developed through a behaviorally based job analysis as described in the procedures section of this article. These first-order factors, defined by a multiple-indicator measurement model, constitute facets of the broader constructs of interest. Each first-order factor or facet denotes a latent variable, which respectively defines specific domains of content and makes up the building blocks of the constructs in our hierarchical model. The identification of these facets with a structural equation model corresponds to the traditional scale construction process; each facet is defined by a unidimensional set of items. In this study, the facets are directly operationalized in terms of a subset of the Likert items administered to 480 team members. Constructs such as Stevens and Campion’s subfactor of collaborative problem solving are operationalized as second-order factors in the model, with the facets as their indicators. Gerbing and Anderson (1984), as well as Rindskopf and Rose (1988), have elaborated on the meaning of these factors and their substantive interpretation.

The hypothesized hierarchical model is intended to provide a meaningful explanation, in terms of latent variables, of the observed correlational relationships among the measured variables. HCFA analyzes the first-order, second-order, and third-order
Figure 1: Validated Hierarchical Factor Structure

NOTE: The covariances among the 5 second-order and 2 third-order KSAs are not represented to ease interpretation. Communication was specified as being measured by a single indicator; therefore, their error variances were fixed at one minus the reliability multiplied by the item variance (Prussia, Kinicki, & Bracker, 1993). Alpha for communication was set at .80.
factors simultaneously. If data on the measured variables are collected from team members, then the HCFA technique can test whether the hypothesized model fits the observed data and can estimate the unknown values of first-order, second-order, and third-order factor loadings, error terms, and correlations in a single analysis. The existence of Stevens and Campion’s (1994) teamwork behavior hierarchy is empirically supported if the hypothesized hierarchical model fits the observed data.

METHOD

PARTICIPANTS

Participants were 480 undergraduate business students (58% were female; mean age was 21 years) in an Organizational Behavior/Human Resources Management course. These 480 participants formed 94 teams. The modal number of participants in each team was 5. There were no missing data.

This sample was chosen for three reasons. First, the ability to work in a team environment is considered a core managerial competency (Allred, Snow, & Miles, 1996). Second, the task assignments were representative of teamwork in organizational settings. Each team was assigned 13 projects; hence, the individuals within each team were required to work together throughout the semester. These team projects represented 20% of an individual’s final grade in the course. The roles of team members were highly interdependent. Each team was identifiable (i.e., had a unique name). Teams had total authority on task planning and assignment of tasks among team members. The decisions made by these teams had consequences for them in terms of grades, both as a team and as individuals within the team. Minimal guidance was given to the teams. Specifically, the only direction provided by the instructor to the teams was the assignments and their completion times.
As these teams were not artificially created for research purposes, they met McGrath’s (1984) definition of natural groups. Moreover, they performed tasks that were organizationally relevant and were autonomous work teams as defined by Guzzo and Dickson (1996):

teams . . . who typically perform highly related or interdependent jobs, who are identified and identifiable as a social unit in an organization, and who are given significant authority and responsibility for many aspects of their work, such as planning, scheduling, assigning tasks to members, and making decisions with economic consequences (usually up to a specific limited value). (p. 324)

Third, team research has generally involved single-part tasks that require individuals to “ideate names or uses or consequences of a thing, or ideate ways to achieve a goal” (Brophy, 1998, p. 213), in short-lived teams in contrived laboratory settings. In this sample, the interactive teams completed a variety of problem-solving tasks over a 13-week period. Moreover, the tasks were completed under constraints that required the active management of time and other resources. We specifically chose problem-solving tasks, as effective problem solving represents a key activity of teams (Katzenbach & Smith, 1994) and also represents a key activity contributing to team effectiveness (Guzzo, 1995).

PROCEDURE AND MEASURES

The key steps in the procedure included the following: assignment of individuals to teams, completion of team problem-solving tasks, evaluation of team tasks, development of the BOS, collection of individual and team performance measures, validation of the BOS, and the completion of HCFA.

Assignment of individuals to teams. The 13-week course in question regularly uses team problem-solving tasks to help students learn and apply theory. All students were randomly assigned
to one of nine sections. In week 1, participants within each section self-selected membership into teams of five to six people.

*Completion of team problem-solving tasks.* As is the case in many organizational settings, tasks assigned to teams required problem identification (e.g., management problems presented in case studies), decision making (e.g., generating options, products, or services), picking evaluation criteria and applying the criteria, seeking additional information (e.g., library research or seeking subject matter experts), critical thinking (e.g., critical evaluation of newspaper articles), building consensus on how best to handle the problem, generating an action plan, implementing the plan, evaluating the outcome and changing decision making and process heuristics in future sessions, and report generation. Minimum guidance was provided on how to complete these problem-solving tasks. Teams were required to complete their tasks within 50-minute sessions. In total, 13 team problem-solving tasks were completed over the 13 weeks.

*Evaluation of the team tasks.* An external judge scored the weekly reports and provided weekly feedback on team performance. This evaluator (who had recently graduated with a Bachelor of Commerce undergraduate degree and was hired by the University as an instructional assistant) was independent of the research group and blind to the study’s purposes. Each week, teams received feedback on the previous week’s report. For each team, feedback consisted of a grade (out of 20 marks) and a one-page written evaluation. The written evaluation and the marking scheme were based equally on the appropriateness of the solution, idea or product, originality, elaboration (amount of detail in the responses), and when appropriate, fluency (total number of relevant responses). A second judge was asked to score a random sample of 40 reports to assess if the criteria were robust across judges. The second judge was a graduate student in psychology who did not observe team sessions and scored reports after week 13 of the study. Analyzing the ratings of the two judges resulted in a $r_{wg}$ (James, Demaree, & Wolf, 1993) of .73 and ICC(2,1) (Shrout & Fleiss, 1979) of .43.
Hence, the initial judge’s ratings appear to be both reliable and valid.

Development of the BOS. The critical incident technique (CIT) is a useful initial step in developing performance assessment tools (Flanagan, 1954; Latham & Wexley, 1994). This technique requires that participants are familiar with the task at hand as well as the behaviors necessary to perform the task effectively. As such, critical incidents were collected after week 11 so that participants would be fully aware of the behaviors necessary to be an effective team member on the problem-solving tasks.

The BOS were developed consistent with the procedures outlined in detail by Latham and Wexley (1994). An abbreviated overview follows. First, study participants (i.e., team members) were given critical incident cards and asked to think about their team experience over the weeks that their team had worked together. They were then asked to recall one example each of effective and ineffective teamwork behavior that they had personally observed take place in their team sessions. Each team member completed at least one card each for effective and ineffective behavior; no more than four cards were collected per team member. Each card asked team members to describe (a) what circumstances led to the incident, (b) what exactly the team member did that was (in)effective, and (c) what the consequences were of the team member’s actions.

Second, two doctoral students (sorters) who were familiar with CIT sorted the 1,356 critical incident cards and developed 14 meaningful clusters. These sorters then gave each cluster a descriptive dimension label (e.g., Performance Management). Next, two other doctoral students (judges) received the same critical incidents in random order and worked together to reclassify the incidents according to the descriptive dimension labels established by the original sorters. The ratio of correctly classified incident/total number of incidents for each cluster was greater than .80, the minimum stated by Latham and Wexley (1994), and thus deemed adequate. The BOS were developed so that the major dimensions, the most frequently occurring incidents, and the incidents judged by the group members as the most important were represented.
Collection of individual and team performance measures. Individual performance measures were collected in week 13. At that time, peers of the same team rated each other’s performance using the BOS. Anonymous peer assessments of team members’ behavior were used, as more than 20 years of scientific research has shown that peer evaluations are reliable and valid (Kane & Lawler, 1978). Moreover, the reliability and validity coefficients of peer ratings are generally significantly higher than supervisory ratings (Kremer, 1990).

Team performance was operationalized as the team’s average score on 13 written reports. There was one report for each exercise. Average scores ranged from 9.65 to 19.69, out of a total possible score of 20 ($SD = 1.42$ and coefficient of variation was .09). The average Cronbach’s alpha coefficient for the 13 submissions was .87.

Validation of the BOS. The BOS developed were validated using both the peer ratings and team performance ratings. In addition, the BOS were cross-validated on a second sample of 176 students in 32 teams that completed team projects over a 13-week period. More details concerning the validation are presented in the Results section.

HCFA. Consistent with the procedure of Gerbing et al. (1994), HFCFA was conducted using the full data sample from the first and cross-validation sample. Additional details are presented in the Results section.

RESULTS

TEAM MEMBER BOS

The resulting BOS contained 14 dimensions, consisting of 46 behavioral items. Table 1 shows the dimensions and items. The BOS dimensions were developed initially using the rational method. That is, the judges grouped the incidents with underlying
processes in mind. Next, each team member rated each of his or her peers using the BOS. Because BOS ratings were based on averaging peer assessments, it was necessary to determine whether peers agreed in their ratings. Agreement was estimated by the average interrater agreement statistic $r_{wg}$ (James et al., 1993), for which values of .70 or better support aggregation (George & Bettenhausen, 1990). The lowest average $r_{wg}$ for BOS dimensions was .74. The range of $r_{wg}$ values across teams for BOS dimensions was .73 to .85. These estimates suggested adequate agreement between peers; consequently, peer assessments were averaged for each participant.

Using these ratings, LISREL 8, maximum-likelihood confirmatory factor analysis (CFA) (Jöreskog & Sörbom, 1996) was conducted to determine whether the item groupings developed by the judges adequately fit the data. CFA revealed adequate fit (Root Mean Square Error of Approximation (RMSEA) = .06, goodness-of-fit index (GFI) = .97, comparative-fit index (CFI) = .98, and normed-fit index (NFI) = .96) (Jöreskog & Sörbom, 1996). A single-factor solution yielded a worse fit to the data ($\Delta$CFI = 5, $\chi^2_{diff} = .197.07, p < .001$).

Factor loadings and a scree plot of eigenvalues from exploratory, maximum likelihood, oblique factor analysis support the confirmatory results. The Kaiser-Meyer-Olkin measure of sampling adequacy was highly appropriate (MSA = .94), and the Bartlett Test of Sphericity classified the data adequately for the analysis (BTS = 13086.15, $p < .001$). For the selection of the number of factors, a typical approach based on roots criterion and the scree test was used. Fourteen factors corresponding to those generated through the rational method were suggested by the exploratory factor analysis. The 14 factors accounted for 75.52% of the variance in the items, and 40 of the 46 items loaded on the expected factors with loadings of greater than .75. Of note is that items forming the three dimensions of “reaction to conflict,” “addresses conflict,” and “averts conflict” loaded on their primary factors with loadings greater than .65 and on the other factors with factor loadings between .55 and .37. This supports a common higher order factor. Also, two items making up the “focus on task-at-hand” factor loaded on their primary factor with a loading greater than .60 but
### TABLE 1: BOS Dimensions and Sample Items

<table>
<thead>
<tr>
<th>Facet</th>
<th>Corresponding Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reaction to conflict</td>
<td>Leaves a conflict unresolved by not saying anything or ignoring some team members (R)</td>
</tr>
<tr>
<td></td>
<td>Leaves a conflict unresolved by leaving the meeting (R)</td>
</tr>
<tr>
<td></td>
<td>Leaves a conflict unresolved by moving on to another topic (R)</td>
</tr>
<tr>
<td>Addresses conflict</td>
<td>Clarifies contentious issues in a conflict</td>
</tr>
<tr>
<td></td>
<td>Politely gives advice in a conflict</td>
</tr>
<tr>
<td></td>
<td>Politely confronts team members on their tardiness</td>
</tr>
<tr>
<td></td>
<td>Provides an alternative solution that is agreeable to other team members when a conflict occurs</td>
</tr>
<tr>
<td>Averts conflict</td>
<td>Resorts to personal attacks when a problem arises (R)</td>
</tr>
<tr>
<td></td>
<td>Tries to calm down team members who are in a conflict</td>
</tr>
<tr>
<td></td>
<td>Takes a stance on an issue and is not willing to budge (R)</td>
</tr>
<tr>
<td>Synthesis of team’s ideas</td>
<td>Builds on the group’s ideas by offering solutions</td>
</tr>
<tr>
<td>Involving others</td>
<td>Clarifies and organizes the group’s ideas</td>
</tr>
<tr>
<td>Participates in problem-solving</td>
<td>Clarifies and explains issues when someone does not understand</td>
</tr>
<tr>
<td></td>
<td>Asks other team members what they think</td>
</tr>
<tr>
<td>Effective communication</td>
<td>Offers ideas</td>
</tr>
<tr>
<td></td>
<td>Asks relevant questions</td>
</tr>
<tr>
<td></td>
<td>Accepts team roles and tasks as required</td>
</tr>
<tr>
<td></td>
<td>Voices unique ideas</td>
</tr>
<tr>
<td>Goal setting/achievement</td>
<td>Dominates the discussion (R)</td>
</tr>
<tr>
<td></td>
<td>Ignores what other team members are saying (R)</td>
</tr>
<tr>
<td></td>
<td>Carefully listens to what others are saying</td>
</tr>
<tr>
<td>Team citizenship</td>
<td>Does not participate in setting team goals (R)</td>
</tr>
<tr>
<td></td>
<td>Participates in developing strategies to achieve team goals</td>
</tr>
<tr>
<td></td>
<td>Volunteers to do things that no one else wants to do</td>
</tr>
<tr>
<td></td>
<td>Keeps working when others quit</td>
</tr>
<tr>
<td></td>
<td>Exercises initiative by acting independently for the benefit of the team (e.g., makes a photocopy for all team members)</td>
</tr>
<tr>
<td></td>
<td>Takes the lead in coming up with ideas</td>
</tr>
<tr>
<td></td>
<td>Seeks information from resources from outside of the team (e.g., books, people, etc.)</td>
</tr>
<tr>
<td>Commitment to team</td>
<td>Misses team meetings (R)</td>
</tr>
<tr>
<td>Focus on task-at-hand</td>
<td>Comes to team meetings late (R)</td>
</tr>
<tr>
<td></td>
<td>Draws team members into off-topic discussions (R)</td>
</tr>
<tr>
<td></td>
<td>Does not try to bring off-topic team members back on topic (R)</td>
</tr>
<tr>
<td></td>
<td>Participates in off-topic conversations (R)</td>
</tr>
<tr>
<td></td>
<td>Draws team members into discussions that are relevant to achieving the goal</td>
</tr>
<tr>
<td></td>
<td>Asks for help in order to get other team members to focus on the goal</td>
</tr>
<tr>
<td></td>
<td>Reminds other team members of the team’s goal</td>
</tr>
<tr>
<td>Preparation for meetings</td>
<td>Does not read the required material prior to team meetings (R)</td>
</tr>
<tr>
<td></td>
<td>Brings the required material to the team meetings</td>
</tr>
</tbody>
</table>
also loaded well on the performance management dimension with a loading of .56 for “does not try to bring off-topic team members back on topic” and .53 for “reminds other team members of the team’s goal.” However, given the judges’ categorization of the factors, we decided to leave them in the “focus on task-at-hand” dimension. Overall, the exploratory factor analysis did not seem to suggest revisions to the hierarchical structure in Figure 1.

*Individual behavior and team performance.* Age, gender, and team size have been found to affect team performance. As such, these variables were controlled for prior to further analysis being conducted. Table 2 reveals correlations between additively aggregated team member behavior and team performance. Each dimension of the BOS achieved a significant correlation ($p < .001$) with the team’s overall performance. The strongest correlation was between “synthesis of team’s ideas” and team performance ($r = .67$, $p < .001$), and the weakest was between “averts conflict” and team performance ($r = .34$, $p < .001$).

A stepwise regression of team performance on the team’s aggregated effectiveness on BOS dimensions was conducted to determine the most predictive combination of BOS dimensions (Table 3) with team performance. The variables that entered and remained in the equation were as follows: (a) synthesis of team’s ideas, (b) participates in problem solving, (c) focus on task-at-hand, (d) involv-
<table>
<thead>
<tr>
<th>Development sample</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Goal setting</td>
<td>1.00</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>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Focus</td>
<td>.55***</td>
<td>1.00</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>
<td></td>
</tr>
<tr>
<td>3. Performance</td>
<td>.50****</td>
<td>.61***</td>
<td>1.00</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>
<tr>
<td>management</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>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Team citizenship</td>
<td>.64****</td>
<td>.57***</td>
<td>.52***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Participation</td>
<td>.68****</td>
<td>.54***</td>
<td>.59***</td>
<td>.71***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Synthesis</td>
<td>.70****</td>
<td>.57***</td>
<td>.64***</td>
<td>.71***</td>
<td>.79***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Team commitment</td>
<td>.23***</td>
<td>.14</td>
<td>.16</td>
<td>.33***</td>
<td>.30***</td>
<td>.32***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Preparation</td>
<td>.37***</td>
<td>.38***</td>
<td>.44***</td>
<td>.51***</td>
<td>.42***</td>
<td>.37***</td>
<td>.29***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Feedback</td>
<td>.64****</td>
<td>.52***</td>
<td>.49***</td>
<td>.65***</td>
<td>.64***</td>
<td>.64***</td>
<td>.20*</td>
<td>.44***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Effective</td>
<td>.47****</td>
<td>.44***</td>
<td>.23**</td>
<td>.32***</td>
<td>.45***</td>
<td>.41***</td>
<td>.12</td>
<td>.36***</td>
<td>.59***</td>
<td>.100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>communication</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>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Involving others</td>
<td>.66****</td>
<td>.54***</td>
<td>.49***</td>
<td>.70***</td>
<td>.82***</td>
<td>.72***</td>
<td>.21*</td>
<td>.35***</td>
<td>.67***</td>
<td>.46***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Reaction to conflict</td>
<td>.46****</td>
<td>.40***</td>
<td>.24**</td>
<td>.38***</td>
<td>.42***</td>
<td>.46***</td>
<td>.12</td>
<td>.26**</td>
<td>.44***</td>
<td>.48***</td>
<td>.38***</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Addresses conflict</td>
<td>.54****</td>
<td>.40***</td>
<td>.50***</td>
<td>.74***</td>
<td>.63***</td>
<td>.65***</td>
<td>.15</td>
<td>.27**</td>
<td>.54***</td>
<td>.23**</td>
<td>.65***</td>
<td>.29***</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Averts conflict</td>
<td>.21*</td>
<td>.35***</td>
<td>.18*</td>
<td>.30***</td>
<td>.27**</td>
<td>.22**</td>
<td>.08</td>
<td>.31***</td>
<td>.42***</td>
<td>.36***</td>
<td>.23**</td>
<td>.34***</td>
<td>.21*</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>15. Team performance*</td>
<td>.55****</td>
<td>.52***</td>
<td>.49***</td>
<td>.52***</td>
<td>.59***</td>
<td>.67***</td>
<td>.35***</td>
<td>.39***</td>
<td>.61***</td>
<td>.48***</td>
<td>.55***</td>
<td>.46***</td>
<td>.46***</td>
<td>.34***</td>
<td>1.00</td>
</tr>
<tr>
<td>Mean</td>
<td>4.12</td>
<td>3.64</td>
<td>3.18</td>
<td>3.54</td>
<td>4.15</td>
<td>3.89</td>
<td>4.53</td>
<td>3.97</td>
<td>4.00</td>
<td>4.15</td>
<td>3.97</td>
<td>4.36</td>
<td>3.25</td>
<td>3.96</td>
<td>15.78</td>
</tr>
<tr>
<td>SD</td>
<td>.56</td>
<td>.29</td>
<td>.41</td>
<td>.34</td>
<td>.31</td>
<td>.38</td>
<td>.30</td>
<td>.45</td>
<td>.27</td>
<td>.31</td>
<td>.40</td>
<td>.33</td>
<td>.41</td>
<td>.31</td>
<td>1.42</td>
</tr>
</tbody>
</table>

**Note:** Significance levels: *p < 0.10, **p < 0.05, ***p < 0.01, ****p < 0.001.
<table>
<thead>
<tr>
<th>Cross-validation sample</th>
<th>1. Goal setting</th>
<th>1.00</th>
</tr>
</thead>
<tbody>
<tr>
<td>2. Focus</td>
<td>.37**</td>
<td>1.00</td>
</tr>
<tr>
<td>3. Performance management</td>
<td>.25**</td>
<td>.46***</td>
</tr>
<tr>
<td>4. Team citizenship</td>
<td>.17</td>
<td>.19*</td>
</tr>
<tr>
<td>5. Participation</td>
<td>.20*</td>
<td>.42***</td>
</tr>
<tr>
<td>6. Synthesis</td>
<td>.26</td>
<td>.38***</td>
</tr>
<tr>
<td>7. Team commitment</td>
<td>.18</td>
<td>.31***</td>
</tr>
<tr>
<td>8. Preparation</td>
<td>.34</td>
<td>.43***</td>
</tr>
<tr>
<td>9. Feedback</td>
<td>.14</td>
<td>.16</td>
</tr>
<tr>
<td>10. Effective communication</td>
<td>.11</td>
<td>.20**</td>
</tr>
<tr>
<td>11. Involving others</td>
<td>.05</td>
<td>.09</td>
</tr>
<tr>
<td>12. Reaction to conflict</td>
<td>.18*</td>
<td>.23**</td>
</tr>
<tr>
<td>13. Addresses conflict</td>
<td>.02</td>
<td>.08</td>
</tr>
<tr>
<td>14. Averts conflict</td>
<td>.10</td>
<td>.00</td>
</tr>
<tr>
<td>15. Team performance b</td>
<td>.44***</td>
<td>.46***</td>
</tr>
<tr>
<td>Mean</td>
<td>3.06</td>
<td>2.96</td>
</tr>
<tr>
<td>SD</td>
<td>.68</td>
<td>.51</td>
</tr>
</tbody>
</table>

a. N = 94 teams.
b. N = 32 teams.
* p < .05. ** p < .01. *** p < .001.
ing others, (e) commitment to the team, (f) goal setting/achievement, (g) performance management, (h) providing/reaction to feedback, and (i) team citizenship. Together these behaviors accounted for 72% of the variation in a team’s performance ($R^2 = .72$, $p < .001$).

Cross-validation. We then cross-validated the results on a second sample. This sample consisted of 176 3rd-year undergraduate business students. This particular sample was chosen because all the students were enrolled in a cooperative degree program and, hence, had full-time work experience. Each student had completed at least three 4-month work placements. Therefore, they may be more representative of potential employees when compared to the participants from the development sample.

These students were randomly assigned to three sections of approximately 42 students each. In week 1 of the 13-week course, students were randomly assigned to 32 teams of four to six individuals (median = 5), as the course contained team assignments requiring critical thinking, analysis, and problem solving. The main tasks focused on identifying and solving organizational problems that they had observed take place in the organizations in which they worked on their last work term. Teams then had to present the team solution to their class and defend their solution. Of a student’s overall course grade, 25% was determined by the team’s output over a 13-week period.

<table>
<thead>
<tr>
<th>Model</th>
<th>Variable Entered</th>
<th>$\Delta R^2$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Synthesis of team’s ideas</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>Participates in problem solving</td>
<td>.03***</td>
<td>.58</td>
</tr>
<tr>
<td>3.</td>
<td>Focus on the task-at-hand</td>
<td>.05***</td>
<td>.63</td>
</tr>
<tr>
<td>4.</td>
<td>Involving others</td>
<td>.02**</td>
<td>.65</td>
</tr>
<tr>
<td>5.</td>
<td>Commitment to the team</td>
<td>.02**</td>
<td>.67</td>
</tr>
<tr>
<td>6.</td>
<td>Goal setting/achievement</td>
<td>.02*</td>
<td>.69</td>
</tr>
<tr>
<td>7.</td>
<td>Performance management</td>
<td>.01*</td>
<td>.70</td>
</tr>
<tr>
<td>8.</td>
<td>Providing/reaction to feedback</td>
<td>.01*</td>
<td>.71</td>
</tr>
<tr>
<td>9.</td>
<td>Team citizenship</td>
<td>.01*</td>
<td>.72</td>
</tr>
</tbody>
</table>

* $p < .05$, ** $p < .01$, *** $p < .001$.  

TABLE 3: Stepwise Regression of Team Performance on the Team’s Average Effectiveness on BOS Dimensions ($N = 94$)
Again, LISREL 8, maximum-likelihood CFA (Jöreskog & Sörbom, 1996) was conducted to determine whether the item groupings developed by the judges in the development sample adequately fit the data in the cross-validation sample. CFA revealed adequate fit (RMSEA = .06, GFI = .91, CFI = .91, NFI = .92). The bottom half of Table 2 reveals the intercorrelations between facets and the significant correlations between facets and team performance (correlation coefficients range from .27 to .51).

**HCFA.** HCFA was also conducted using LISREL, with maximum-likelihood estimates derived from a covariance matrix. Upon obtaining a successful first-order model, the operationalization of the constructs as second-order factors could proceed, followed by the third-order analysis. Once the construct structure was formed, the structural relations among the constructs could be modeled. Successively embedding each model within the succeeding model led to the specification of the full model that simultaneously analyzed the relationships between all manifest and latent variables. Gerbing et al. (1994) describe in detail how this is accomplished. The magnitude of relations between the third-order, second-order factor, and the facets was initially judged by the factor loadings generated by HCFA.

The standardized first-, second-, and third-order factor loadings shown in Figure 1 were all statistically significant (t values varied between 8.31 and 11.74) and, thus, important to the hypothesized hierarchical model. Consistent with a hierarchical representation, the teamwork third-order constructs (i.e., self-management and interpersonal skills) were significantly related to all five of Stevens and Campion’s subcategories of conflict resolution, collaborative problem solving, communication, goal setting and performance management, as well as planning and task coordination. The values of these second-order factor loadings ranged from .47 to .84, and the values of third-order factor loadings ranged from .31 to .51. The communication factor obtained the highest third-order factor loading (Gamma = .48, p < .05) on interpersonal KSAs, meaning that a standard deviation change in this higher order construct was associated with a .48 standard deviation change in interpersonal KSAs.
The conflict resolution and collaborative problem-solving factors were less strongly related to interpersonal KSAs, but the relationship was still significant (Gamma = .45, \( p < .05 \), and Gamma = .31, \( p < .05 \), respectively). The goal setting and performance management factor obtained the highest third-order factor loading (Gamma = .51, \( p < .05 \)) on self-management KSAs. The planning and task coordination factor was less strongly related to self-management KSAs (Gamma = .49, \( p < .05 \)).

Validation of the complete model was demonstrated by the Relative Noncentrality Index (RNI) (McDonald & Marsh, 1990), which was .92, indicating acceptable fit. CFI (Bentler, 1990)—a bounded version of RNI in which values larger than 1 are truncated to 1 and values less than 0 are truncated to 0—was also adequate at .92. These measures were chosen because Gerbing et al. (1994) argue that these are the best measures of fit for HCFA when there are many indicators.¹

**DISCUSSION**

This study was designed to identify the key behaviors necessary for effective individual performance on team problem-solving tasks and to examine the relationship between these behaviors and team performance on such tasks. The HCFA provides a means to simultaneously incorporate a full item analysis in the same structural equation model that also operationalizes the broadly defined constructs defined by Stevens and Campion (1994). As such, the practical and theoretical implications of this work are at least six-fold.

First, this study provides one of the first typologies of performance-relevant, individual team member behaviors based on a systematic job analysis of observed team member behavior and embeds that typology within an existing hierarchy of broader constructs. In doing so, it provides support for Stevens and Campion’s (1994) hierarchy. As one moves up the factor hierarchy, one should expect that broader constructs are more likely to transpose to other tasks. For instance, “synthesis of team’s ideas” (facet) may be
important only for problem-solving teams, but “interpersonal skills” (third-order factor) is a more generic factor or competency that would be important for all teams. As shown in Table 4, the pres-
tent typology provides empirical support for previous typologies, as it contains most of the behaviors identified in these previous studies and, in addition, identifies two dimensions not previously identified (i.e., team citizenship, commitment to the team).

However, the present BOS did not capture a key criterion found in some other typologies, namely, leadership. A potential explanation for this omission may lie in the nature of our teams, namely, that they were self-directed, initially leaderless teams. A closer examination of the BOS indicates that many of the factors in Figure 1 consist of behaviors associated with leadership (e.g., performance management, conflict resolution, and task coordination). Taggar, Hackett, and Saha (1999) suggest that a team leader is simply an individual who is more likely to exhibit certain effective team member behaviors attributed to leadership than are others. As such, future research should assess whether these BOS generalize to teams that are not self-directed and whether leadership is a separate factor in teams with leaders or if it is better represented as a combination of BOS.

Second, in response to the call of Cannon-Bowers et al. (1995), this study is an attempt to build a consensus in the literature. Rather than researchers working individually, the field now needs to build agreement on what constitutes effective teamwork behavior. As described previously, this study attempts to facilitate this consensus by examining the relationship between KSAs developed through a CIT job analysis, and presented in BOS, and the previous ones, in particular the KSAs of Stevens and Campion (1994) that were inferred through a literature review.

Third, this study adds to the literature through the validation of the present BOS. To our knowledge, this is the first time that a study has demonstrated a statistically significant relationship between a comprehensive behaviorally specific typology of team member behavior and team performance on team problem-solving tasks. In fact, as previously discussed, Miller (1999) failed to detect a significant relationship between the frequently cited Stevens and Campion’s (1994) teamwork KSAs and team performance on problem-solving tasks.
Fourth, from a practical perspective, behavioral criteria can be used to show people what they need to stop, start, or continue doing to be effective performers (Latham & Wexley, 1994). Hence, following the advise of Stevens and Campion (1994), the present BOS can be used as the basis of performance management and development programs designed to improve and develop individual team member behavior on team problem-solving tasks. Moreover, based on a wealth of data supporting the use of goals and BOS instruments (see discussion in Locke & Latham, 1990), as well as the findings of Brown and Latham (1999), who demonstrated that goal setting using BOS that included KSAs needed for effective teamwork increased individual performance, the present BOS can be used in conjunction with goal setting to help people develop effective teamwork behaviors.

Fifth, these BOS can be used for recruitment purposes in that they can help determine the best combination of behaviors for predicting team performance on problem-solving tasks. The importance of problem-solving tasks was highlighted previously. Our results suggest that when staffing intact work teams that complete a variety of tasks, practitioners should select people who synthesize the team’s ideas, participate in team problem solving, focus on the task at hand, involve others, are committed to the team, participate in goal setting/achievement, manage the performance of other team members, provide feedback and react well to feedback, and are good team citizens. Hence, these BOS dimensions may be incorporated into behaviorally based selection tools such as the situational interview, which is often used in conjunction with BOS (Latham & Sue-Chan, 1996) to recruit people who will be effective team members for these tasks.

Sixth, the BOS dimensions can also be used to determine the role a new or existing member should assume within a team. The trend toward team-based organizations is largely based on the belief that teams should be composed of people with complementary skills so that one member’s weakness is counterbalanced by another member’s strength (Katzenbach & Smith, 1994). Hence, the present BOS dimensions can be used to assess what skill gaps exist in the
team and then used to recruit a team member who has the needed skill set. For instance, a person who exhibits the behaviors captured under the “performance management” and “goal setting/achievement” BOS dimensions may act well in an initiating structure kind of leadership role (cf. Cohen, Chang, & Ledford, 1997). If a team needed a gatekeeper (someone who facilitates the participation of others in the team), then the team should seek someone who exhibits the behaviors found under the “involving others” BOS category. Similarly, the usefulness of selection criteria in predicting team performance may vary in response to a team’s mandate (Hyatt & Ruddy, 1997). One may work backward from the team task to determine the essential team behaviors. For instance, a management team established to negotiate a collective agreement might benefit from people who exhibit effective conflict-related BOS behaviors (avert conflict, react well if a conflict does occur, and forward a strategy to address conflict). Similarly, if an existing team lacks a person to take on a specific role, then the desired behaviors may be developed in its members.

The potential limitations of this study are four-fold. First, the focus of this study was on intrateam behavior. Future research is required on interteam behavior and organizational support systems, building on the contribution of Hyatt and Ruddy (1997), who recognize the importance of considering teams embedded in larger organizational contexts.

Second, this study attempted to approximate genuine work environments while benefiting from a large sample with equivalent teamwork experiences and resource constraints. The teams appeared similar to “real” work teams in task interdependence—this was supported by the emergence of “performance management,” “participates in problem solving,” “synthesis of team’s ideas,” and “involving others” dimensions. Although there is some support for using students as subjects (Greenberg, 1987), future studies of functioning intact work teams within firms are needed to establish generalizability.

Third, the BOS developed in this study has been validated on team problem-solving tasks. The extent to which this BOS can be an effective instrument for other team tasks is unknown. Future
research should examine the extent to which the BOS can generalize to other team tasks.

Fourth, as pointed out by an anonymous reviewer, two of the first-order factors (performance management and effective communication) may require relabeling to improve clarity. Specifically, the first-order factor of performance management fell under the second-order factor of planning/task coordination rather than the second-order factor of goal setting/performance management, and the first-order factor of effective communication fell under the second-order factor of communication. Although prior to this point in the article we were hesitant to change the labels given to these factors, as naming of the dimensions by sorters is a key element of the BOS development process (see Latham & Wexley, 1994), we now suggest alternative names in parentheses in Figure 1. Specifically, we suggest the alternatives of “process management” and “active listening” for “performance management” and “effective communication,” respectively.

It has been more than 6 years since Stevens and Campion (1994) proposed their typology of teamwork behavior. Since then, their article has been widely cited, however, little progress has been made in elaborating and building on their typology. They noted the need to determine “how individual teamwork performance is linked to the performance of the entire team” (p. 522). This study supports Stevens and Campion’s typology and provides elaboration. Further, we have shown and replicated the finding that aggregated individual performance, as assessed by the BOS developed here, significantly impacts overall team performance.

NOTE

1. Gerbing, Hamilton, and Freeman (1994) note that a standard fit index for structural equation models has traditionally been Jöreskog and Sörbom’s (1996) goodness-of-fit index (GFI), but deficiencies in this index have led to the development of indices based on the noncentral chi-square distribution, such as RNI. Consistent with previous work in which GFI was low even for correctly specified models with many manifest variables (Anderson & Gerbing, 1984), the GFI of this model was only .85, which is low when compared to the acceptable .90 value of RNI.
REFERENCES


Simon Taggar received his Ph.D. from McMaster University. He is an assistant professor in human resource management at York University. His research interests include team processes, with an emphasis on team leadership and creativity.

Travor C. Brown received his Ph.D. from the University of Toronto. He is an assistant professor in human resource management and labor relations at Memorial University. His research interests include team processes, with an emphasis on goal setting.
THE PROVISION OF EFFORT IN SELF-DESIGNING WORK GROUPS
The Case of Collaborative Research

NATHAN BENNETT
Georgia Institute of Technology

ROLAND E. KIDWELL, JR.
Niagara University

Teams of academic coauthors can be conceptualized as self-designing work groups, an infrequently studied but increasingly prevalent group structure. This research note considers issues surrounding how management scholars form collaborative teams, provide effort toward completion of research projects, evaluate colleagues' efforts, and decide whether to pursue further collaborative opportunities with them. The findings indicate that withholding effort occurs in self-designing groups, such as research collaborations, and that the emotional bonds that group members form with colleagues play a key role in whether they decide to work together again, as well as in how they react to perceptions that a coauthor withheld effort.

A review of the literature on work groups in organizations reveals a paucity of research focused on what Hackman (1987) termed self-designing work groups. Such groups cooperatively determine their membership, manage their own activities, perform their own tasks, and develop their own norms to guide decision making. Examples of such groups provided by Hackman include top management...
groups, boards of directors, and mature autonomous work teams. It is particularly important to conduct research on various types of self-designing work groups because team structures are used in almost half of all organizations and the use of ongoing project teams is becoming more frequent (Devine, Clayton, Philips, Dunford, & Melner, 1999). To this point, research on self-designing work groups has generally focused on top management teams and has explored issues such as conflict (e.g., Amason, 1996) and heterogeneity (e.g., Hambrick, Cho, & Chen, 1996).

Clearly, many important questions about self-designing work groups have yet to be addressed. One notable issue concerns factors that influence group members’ decisions about their provision of effort toward accomplishment of group goals. Although this question has been the focus of research in other types of groups (e.g., George, 1995; Miles & Greenberg, 1993; Wagner, 1995), unique characteristics of self-designing groups suggest that extant research may not generalize. The purpose of this article is to explicitly focus on those factors associated with the provision of effort in self-designing work groups. We do so by examining groups of management scholars whose interdependent task is the production of an academic manuscript. Readers familiar with such groups will see that they cooperatively determine membership and manage their own activities.

Among academicians, it is understood that publishing in high-caliber, peer-reviewed journals is critically important. Academic departments benefit from enhanced reputations when their faculty members publish in respected outlets (Stahl, Leap, & Wei, 1988), and publishing has personal importance because it is the basis for many personal and professional rewards (Cole & Cole, 1967; Gomez-Mejia & Balkin, 1992; Park & Gordon, 1996). To accomplish publishing goals, scholars often design and participate in cooperative groups. The long-observed and well-documented trend toward collaborative publication (Broad, 1981; Floyd, Schroeder, & Finn, 1994; Over, 1982; “Really big science,” 1995), as well as the central role that publication records play in personnel decision making within the university (e.g., Gomez-Mejia & Balkin, 1992), suggests that this performance context is worthy of
attention. Through enhancing our understanding of the process, we gain knowledge of how those who participate in such groups form them, work toward achieving their goals, and evaluate the contributions of fellow team members.

PROVIDING EFFORT IN INTERDEPENDENT TASKS

Kidwell and Bennett (1993) presented a model, based on Knoke (1990), which suggests that individuals working in groups may or may not provide full effort for three general reasons: as a rational choice, in conformance to group norms, or to express affective bonding with coworkers. These three reasons are detailed below and then applied to this study.

RATIONAL CHOICE

The rational choice perspective holds that individuals in work groups decide whether to provide effort based on cost-benefit calculations. The economics literature argues that employees have an increased tendency to supply less effort (i.e., shirk) (Leibowitz & Tollison, 1980) when they can opportunistically take advantage of monitoring difficulties (Alchian & Demsetz, 1972). For example, as group size increases, the contribution of individual members tends to decrease because they believe they can hide in the crowd (e.g., Latané, Williams, & Harkins, 1979). In addition, in unstructured or ambiguous tasks that require greater interdependence to complete, there may be a tendency for effort to be withheld by individuals in work groups because monitoring of effort becomes more difficult as task performance becomes less discrete (Jones, 1984).

When rational choice is considered in the context of a jointly authored academic article, coauthors may attempt to calculate maximum personal utility, providing the minimum effort that would get the paper published. Or, coauthors may determine that they should provide as much effort as possible because the benefits (promotion and tenure, enhanced reputation) outweigh such costs
as compensatory activity for a colleague’s perceived lack of effort (cf. Williams & Karau, 1991). Selection of coauthors during design might be part of a cost-benefit analysis for those who consciously attempt to withhold effort or shift responsibilities to others. Nontenured faculty who need publications to be promoted might seek a self-designing group that includes other faculty as coauthor experts who will carry more of the load as well as assist them in networking activities. Senior faculty may team with junior faculty who can be relied on to provide high effort to solidify their positions in the field through publications, or to otherwise ingratiate themselves. Faculty of all ranks might identify promising Ph.D. students who would provide large amounts of effort on a project. Coauthors may seek to increase the number of collaborators, which would lessen visibility of their own efforts; this strategy contains risks in that too many free riders could sink the entire project. Whereas there are clearly a number of factors that likely influence rational choice motives in regard to self-designing work groups, a starting point for our consideration lies in the following prediction:

Hypothesis 1: As size of the self-designing group increases, withholding effort by individual members tends to increase.

NORMATIVE CONFORMITY

The normative conformity perspective suggests that individuals make choices about withholding effort based on conforming to principles of acceptable behavior. For example, Akerlof (1982, 1984) proposed that norms defining a fair day’s work play a major role in the effort workers are willing to provide. The self-interest of the rational choice perspective is tempered by the idea of a ‘norm of fair dealing’ (Stroebe & Frey, 1982, p. 127) to which individuals comply as a matter of reciprocity toward others. Compliance norms (Heckathorn, 1990) develop within the work group, and these important values take on significance as a “social contract” that may rival rational calculation of costs and benefits. Another normative effect on withholding effort may occur when individuals believe that coworkers will withhold effort and allow them to
shoulder most of the work (Jackson & Harkins, 1985; Schnake, 1991). Instead of compensating for the lower effort levels of coworkers, these individuals reduce their own efforts to avoid being played for suckers.

There are a variety of ways that normative conformity could play a role in a collaborator’s decision to withhold effort on a joint publication. For example, it is a norm within many disciplines that the order of authorship reflects contribution to the manuscript. Thus, we would expect that lower positioned authors provide less effort, relative to higher positioned authors (Floyd et al., 1994). In addition, norms of reciprocity might indicate that on a previous collaboration, Author A did the bulk of the work and Author B withheld effort; on a current collaboration, Author B is expected to shoulder a greater burden of work. In all, the normative conformity perspective suggests the following:

**Hypothesis 2:** Members of self-designing research groups are expected to conform to prevailing norms by contributing greater effort levels based on author order.

**AFFECTIVE BONDING**

Finally, the affective bonding perspective suggests that individuals provide or withhold effort based on their emotional attachments to others. These attachments occur as part of the individual’s identification with other members and with the group. “The resulting sense of ‘oneness’ between person and group strengthens the member’s motives for contributing personal resources to the organization” (Knoke, 1990, p. 42). How much cooperation occurs within a group may be determined by whether the group members plan to work together in the future or already know they will work together again (Axelrod, 1984; Spicer, 1985).

Tying affective bonding to academic collaboration and withholding effort would involve considering personal relationships among coauthors (Floyd et al., 1994). How well coauthors regard each other, whether they identify with coauthors due to demographic similarity, whether personal relationships are important
among them, and whether they believe they will work together again might be expected to affect how much effort is provided or withheld within the conceptual framework of affective bonding. The affective bonding perspective suggests the following hypothesis:

Hypothesis 3: The greater degree that members of the self-designing group like each other, the less likely the members are to withhold effort.

METHOD

SAMPLE

We mailed questionnaires to authors who had published a paper with at least one and as many as four other individuals during 1993, 1994, or 1995 in the Academy of Management Journal or the Academy of Management Review. We did not include any paper with six or more authors because there were so few such articles in the time span considered. This raised concerns that it might not be possible to guarantee anonymity to these respondents. Each questionnaire was coded so that coauthors’ responses could be matched with one another; no key was kept to tie any code number to an article. Further, we were blind to this coding process; there was no way for us to link any author or article to a returned questionnaire. The focus of the questionnaire was on the collaborative effort involved in producing the article; individuals who published more than one article in those journals during that time period received a questionnaire for each article.

Questionnaires were mailed to 418 coauthors of 197 articles. Of the 418 questionnaires mailed out, we received 241 usable individual responses, an overall response rate of 57.7%. Although we received at least one response for 76.7% of the published articles, our analysis considered only those articles where at least two coauthors responded. The final sample size was 165. Within that group, 40% of the respondents were women. At the outset of the focal col-
laboration, 22% were Ph.D. students, 33% assistant professors, 22% associate professors, and 23% full professors. Forty-three percent of the respondents were first author on the collaboration, 38% second author, 17% third author, and 2% fourth author.

MEASURES

A number of measures were included that allowed coauthors to evaluate one another on various aspects of the collaboration. First, to tap the social contract that existed among the coauthors, respondents were asked to imagine that 100 points represented the amount of effort necessary to bring the project to fruition. Then, respondents allocated the 100 points across the coauthors to reflect the contribution each was expected to make at the outset of the collaboration. Respondents then were asked to recall the actual contribution of each coauthor to the project. We computed an effort variable where actual contribution was subtracted from the intended contribution: A positive value indicates the author provided less than intended, a negative score indicates the author provided more than intended. Deviations from a score of 0 (zero) reflect a deviation from the social contract. The score on this variable assigned to each author is the average of their coauthor’s responses. For example, on a paper with three authors, A, B, and C, author A received the mean of the perceptions of B and C with regard to his or her effort, and so on.

Second, respondents answered a series of questions concerning each coauthor. These questions were answered using a Likert-type scale and are coded such that a low score represents a high amount of the focal construct. Respondents completed a four-item measure of coauthor liking (adapted from Wayne & Ferris, 1990; $\alpha = .91$). Sample items include “I would like to spend more time with this person” and “I regard this person as a good friend.” In addition, two original measures were included. A five-item scale ($\alpha = .93$) asked each respondent to make a coauthor evaluation of the quality of each coauthor’s contribution to the paper. Items gauged the degree to which each author’s contributions to the manuscript “were of high quality,” “were completed in a timely fashion,” and “met my
expectations in regard to quality.” A seven-item scale (α = .90) assessed coauthor satisfaction. Sample items include “Working with this individual was a positive experience,” “This person took advantage by withholding effort on this project” (reverse coded), and “I would not accept an opportunity to work with this person in the future” (reverse coded). Again, individuals were assigned values on these measures that were the mean of their coauthors’ responses.

A number of variables were used to describe the context in which the collaboration occurred. Respondents were asked to indicate their rank at the outset of the collaboration. This measure is coded using five categories, where 1 indicates the respondent was a Ph.D. student and 5 indicates the respondent was a full professor. Gender is coded so that 1 represents women and 2 represents men. A three-category variable was used to indicate when the coauthors determined author order on the publication. A score of 1 indicates that the decision was made at the outset of the collaboration, 2 indicates the decision was made during the development of the manuscript, and 3 indicates the decision was made right before the submission of the manuscript, based on contribution. A three-category measure was used to tap liking at outset of the collaboration. Here, a low score indicates greater liking. Respondents were asked (a) whether or not they felt the focal paper was one of their better publications, (b) if they had a previous collaboration experience with any of these coauthors, and (c) if, at the outset of this project, they anticipated future collaboration with any of these coauthors. Each of these variables is coded (1, 2) such that 1 indicates an affirmative response.

Finally, respondents answered a series of items that were developed based on previously conducted interviews with management faculty that assessed (a) the degree to which each coauthor should be credited for various sorts of contributions to the manuscript (e.g., methodological competence, expertise in the topic area), and (b) how influential various factors were in deciding authorship order (e.g., alphabetical, writing of the first draft). Responses to these items were made using 5-point Likert-type scales. In the case
of the former measure, respondents were assigned the mean of the evaluations by their coauthors.

RESULTS

At the outset, we posed three hypotheses to guide this study. Hypothesis 1, which predicted decreased effort levels as size of the group increased, was not supported, perhaps in part because members’ effort levels in their self-designing groups were identifiable and specific. Hypothesis 2, which predicted greater effort levels on the part of the first author due to established norms, was supported. Hypothesis 3, which predicted that the degree of liking would lessen the likelihood of effort decrement, was not supported.

Despite modest support for the hypotheses, a number of other findings provide important insight into the operation of self-designing work groups. Respondents were asked to indicate on a 5-point scale how influential eight possible decision rules for determining authorship order had been in their collaboration. The two criteria receiving the strongest responses were “writing the first draft” ($M = 1.82, SD = 1.27$) and “effort given to the project” ($M = 1.93, SD = 1.18$). Factors such as “contributing a data set” ($M = 2.97, SD = 1.63$) and “methodological competence” ($M = 3.39, SD = 1.45$) were of moderate importance, and others such as alphabetical order ($M = 4.34, SD = 1.12$) were evaluated as unimportant. It is somewhat interesting to note that effort given to the project was so highly regarded as a means for determining authorship when 50% of the respondents also indicated that author order was decided before the collaboration began; only 14% of the respondents indicated that author order was determined after the manuscript was completed.

Respondents were also asked to report the degree to which each author should receive credit for contributing (a) writing expertise, (b) topic expertise, and where relevant, (c) data collection, (d) a data set, (e) methodological expertise, and (f) statistical expertise. The results suggest that contributions of authors two, three, and
four do not differ significantly from one another. There were no remarkable differences for these authors across the various forms of contribution, but there were some differences noted in the contributions of the first author as compared to the others. Specifically, first authors were more often credited for data collection ($t = 2.53, p < .05$) and for being a “topic expert” ($t = 2.07, p < .05$) than were subsequent authors.

Table 1 contains the correlation matrix and descriptive statistics for the variables used in our multivariate analyses. A number of interesting relationships can be noted. First, the four coauthor evaluations that serve as dependent variables in subsequent analyses are moderately intercorrelated, with the exception of the relationship between the effort measure and the liking measure. Among the other variables, the correlations indicate that as the number of authors goes up, author evaluations of the paper’s quality go down. There is a significant correlation between position in authorship and coauthor evaluations of effort; first authors received higher effort scores than subsequent authors. We followed this up with an ANOVA that indicates second authors were rated the lowest on this effort measure—they had, on average, the greatest shortfall between effort expected and effort actually provided ($F_{2,154} = 10.17, p < .001$). There are also moderate intercorrelations between previous collaborations with coauthors, liking of coauthors, and plans for future collaboration at the outset of the project.

In these data, gender was consistently correlated with lower evaluations by coauthors with women receiving lower evaluations than men, no matter the gender of the rater. More positive evaluations, even in the case of author liking, were made of authors who were more highly placed in terms of author order. Previous collaboration with some or all of the coauthors was associated with better liking of each coauthor at the conclusion of the project. Respondents who generally liked their coauthors at the outset of the project reported greater satisfaction with each coauthor at the conclusion of the project as well.

To further consider these intercorrelations, a series of multiple regression analyses were conducted. Based on the exploratory nature of this article, we selected variables for this analysis based


<table>
<thead>
<tr>
<th></th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Coauthor liking</td>
<td>1.81</td>
<td>.81</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>
<tr>
<td>2. Coauthor evaluation</td>
<td>1.57</td>
<td>.80</td>
<td>.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Coauthor satisfaction</td>
<td>1.72</td>
<td>.78</td>
<td>.56</td>
<td>.71</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Effort</td>
<td>.26</td>
<td>7.1</td>
<td>.15</td>
<td>.37</td>
<td>.40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Number of authors</td>
<td>2.63</td>
<td>.66</td>
<td>.32</td>
<td>.06</td>
<td>.03</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Gender</td>
<td>1.60</td>
<td>.49</td>
<td>-.22</td>
<td>-.16</td>
<td>-.14</td>
<td>-.17</td>
<td>-.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Rank</td>
<td>.35</td>
<td>.48</td>
<td>.07</td>
<td>-.07</td>
<td>.05</td>
<td>.11</td>
<td>-.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Author order</td>
<td>1.78</td>
<td>.79</td>
<td>.23</td>
<td>.18</td>
<td>.30</td>
<td>.26</td>
<td>.30</td>
<td>.05</td>
<td>-.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Better publication</td>
<td>1.11</td>
<td>.32</td>
<td>.15</td>
<td>.00</td>
<td>-.02</td>
<td>.20</td>
<td>-.13</td>
<td>-.14</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Previous collaboration</td>
<td>1.38</td>
<td>.49</td>
<td>.22</td>
<td>.07</td>
<td>.05</td>
<td>-.01</td>
<td>.04</td>
<td>-.08</td>
<td>.11</td>
<td>-.05</td>
<td>-.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Liking at outset</td>
<td>1.03</td>
<td>.18</td>
<td>.11</td>
<td>.07</td>
<td>.20</td>
<td>.08</td>
<td>.00</td>
<td>-.16</td>
<td>.09</td>
<td>.10</td>
<td>.06</td>
<td>.23</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Future collaboration</td>
<td>1.08</td>
<td>.27</td>
<td>.14</td>
<td>.03</td>
<td>.08</td>
<td>-.03</td>
<td>.12</td>
<td>-.01</td>
<td>-.01</td>
<td>.05</td>
<td>.06</td>
<td>.32</td>
<td>.23</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE:** $N = 152$. A correlation of .16 is significant with $p < .05$, of .21 with $p < .01$, and of .26 with $p < .001$. 
on our expectations and the correlational results. Results of these analyses are reported in Table 2. The results are fairly consistent across the four dependent variables. Specifically, women and individuals who were lower in terms of author order received poorer evaluations from their coauthors. In addition, individuals who had not previously worked with a particular coauthor and authors on projects with greater numbers of authors were less well-liked by their group members. Because we were particularly surprised by the finding concerning women and their evaluations by coauthors, we repeated these analyses in two separate ways. First, we included as a control variable a measure of the proportion of authors who were women. Second, we conducted the analysis for only those instances where women were the evaluators. In each case, results of the regression analyses were the same. We then conducted a MANOVA to determine if the regression results would hold when intercorrelations between dependent measures were considered. The results suggest they do (Wilks’ $\Lambda = .89, F_{4,138} = 3.97, p < .01$).

Finally, we examined the correlations between respondent evaluations of each of their coauthors and their reports of whether or not they have worked with them since this collaboration. The only evaluation associated with whether or not the two collaborated again was liking; the correlations between liking and a consequent col-

<table>
<thead>
<tr>
<th>Coauthor Liking</th>
<th>Coauthor Evaluation</th>
<th>Coauthor Satisfaction</th>
<th>Effort</th>
</tr>
</thead>
<tbody>
<tr>
<td>β</td>
<td>β</td>
<td>β</td>
<td>β</td>
</tr>
<tr>
<td>Number of authors</td>
<td>.25**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Position in authorship</td>
<td>.17*</td>
<td>.19*</td>
<td>.31***</td>
</tr>
<tr>
<td>Previous collaboration</td>
<td>.20**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Liking at outset</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Better publication</td>
<td>−.20**</td>
<td>−.17*</td>
<td>−.16*</td>
</tr>
<tr>
<td>Gender</td>
<td>−.20**</td>
<td>−.17*</td>
<td>−.16*</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.21</td>
<td>.06</td>
<td>.11</td>
</tr>
<tr>
<td>$F$</td>
<td>9.51***</td>
<td>4.79***</td>
<td>9.50***</td>
</tr>
</tbody>
</table>

NOTE: $N = 150$.  
* $p < .05$. ** $p < .01$. *** $p < .001$.  

### TABLE 2: Results of Multiple Regression Analyses
laboration ranged from .28 to .59 across the up to four authors considered. There was no relationship between evaluations of the coauthors’ contribution to the collaboration and having subsequently worked with them on a project.

**DISCUSSION**

The results of this study provided only modest support for the hypotheses, but there were several interesting findings with regard to self-designing work groups in the context of academic collaboration that could provide guidance for future research into these types of groups. In particular, the results indicated that the norms and emotional bonds that form among members of these self-designing groups—in particular whether the members like each other—sufficiently mitigated adverse actions toward group members who withheld effort.

An important issue in this study was whether withholding effort can be attributed to rational choice, normative conformity, and affective bonding factors. Here, our goal was to apply a theoretical framework (Kidwell & Bennett, 1993; Knoke, 1990) to the process of self-designing groups. At the outset, we believed individuals might be motivated to withhold effort for rational reasons, in conformance to norms, and because they have strong bonds with others in the work group. In this study, the importance of affective bonding and the liking that coauthors had for each other emerged as paramount, particularly in how withholding effort is received when it occurs in a collaborative project, just not as we had predicted.

The results also underscored the importance of equity and reciprocity norms in joint academic efforts. About half of the respondents reported that they had negotiated an informal contract deciding authorship order before the project began. This result in combination with the finding that “effort given to the project” is an important determinant of authorship order indicates that contributions to the project were to be governed by normative expectations (i.e., first authors do more, as predicted in Hypothesis 2). Most individuals then were perceived to have provided strong effort toward
completion of the project. In addition, discrepancies occurred between expected effort negotiated when the group was formed and the evaluation of effort actually provided during the project. Generally, first authors contributed more effort than expected, whereas second—and to a lesser extent, third—authors provided less than expected effort.

When effort given was less than expected, it did not seem as critical in these cases from a post hoc perspective. Perhaps this is due to the fact that the results of the self-designing group were so successful. Those who were perceived to have given less than full effort on a publication that resulted in an *Academy of Management Journal* or *Academy of Management Review* article were not punished for their disregard of the normative expectations. In these data, there was no relationship between a coauthor’s evaluation of a colleague’s efforts and whether they worked together again. The only relevant consideration in whether or not coauthors worked together again appeared to be whether they liked each other.

It is important to reiterate that the study focused respondents on collaborative efforts that were a “success,” in that a high-quality product (i.e., publication in a well-regarded, peer-reviewed journal) resulted. Even in these successful collaborations, some discrepancies between what others expected of one another and what actually was forthcoming were noted. Future research should consider the way dysfunctional group member behavior, including withholding effort, might operate in projects that fail to develop to their potential.

Affective bonds played a crucial role in deciding whom to work with: an expectation that high effort would be provided on the current project based on past experiences or interpersonal relationships. Whether coauthors liked each other was often of greater significance than the performance of individuals on a given project. However, the results contradict the prediction made in Hypothesis 3 that the degree of liking would keep effort levels up within the group.

In retrospect, elements of a motivation model proposed for traditional work groups (e.g., Kidwell & Bennett, 1993) do not appear appropriate in this case. Self-designing groups differ from work
groups in that they form voluntarily. In this study, individuals contributed effort based on rational and normative considerations, as the model would predict, but they also evaluated others’ shortcomings less severely. The unforced nature of the self-designing team underscores the importance of affective bonds among the team members: if potential group members do not like each other, they may not form a team, or they are in a better position to exit than a work group member who may not get along with coworkers.

Clues to the importance of liking in the evaluation of coauthors could be gleaned from an examination of the theoretical roles of interpersonal attraction and leader-member exchange quality on performance appraisals and working relationships. A long stream of research stemming from Byrne (1971) indicates that perceptual similarity is positively related to a manager’s evaluation of subordinates and vice versa (e.g., Pulakos & Wexley, 1983), and from Graen (1976) that a high-quality exchange relationship can be related to high-quality performance. For example, research has offered empirical support for a linkage between liking and performance ratings (Wayne & Ferris, 1990). Whereas these relationships deal with manager/subordinate relationships, future research on collaborative authorship might consider interpersonal attraction or exchange quality among coauthors as theoretical frameworks.

The findings concerning the relationship between gender and evaluations were interesting. Controlling for a number of factors, women received lower evaluations from their colleagues—whether those colleagues were men or women. First, it should be noted that the evaluations were generally favorable; that is, a lower evaluation is still, in most cases, a positive evaluation. At the same time, the consistency of this finding across four dependent measures, and in the presence of a number of control variables, suggests that something may be operating that is worthy of further consideration. There are a number of explanations that might account for the finding that men evaluate women less favorably (e.g., Pulakos & Wexley, 1983). In this study, women also rated women coauthors less favorably than their male coauthors did. This finding merits additional research in the context of self-designing work groups.
There are several other avenues for future research that emerged from this study. First, it would be interesting to expand the examination of coauthors as self-designing groups in a number of ways. For example, research that investigates coauthor behavior in failed collaborative efforts would provide a useful companion to our results concerning successes. Second, research that further explores the dynamics involved in the negotiation of the social contract used to structure a self-designing group would be useful. Such research could consider status differentials among collaborators and how the power held by each party influences the distribution of effort specified in the social contract. Third, in semi-structured interviews conducted prior to data collection, we repeatedly heard accounts of individuals who withheld effort in other areas of professional importance (e.g., team teaching, committee work, service work, professional activities). This leads us to suggest that investigations of withholding effort in other types of academic self-designing groups would be viable.

This study suggests that collective action models provide a level of insight into the functioning of self-designing work groups. That is, members come together for rational reasons, engaging in cost-benefit analyses by deciding which projects they wish to participate in and with whom. Members develop normative expectations of equity in effort and reciprocity by informally negotiating who does what in the collaborative process and agreeing on a first author who is to carry the brunt of the load. Most important, the members’ emotional bonds with other individuals play a strong role in their decision as to whom to work with on a repeating basis. The personal regard they have for fellow group members is integral to how they evaluate their collaborators’ contribution toward the project’s completion.

REFERENCES


Nathan Bennett is a professor of management and associate dean in the DuPree College of Management at the Georgia Institute of Technology. His current research interests include justice in organizations, work group performance, and multilevel models of organizational behavior.

Roland E. Kidwell, Jr., is an assistant professor of management in the Department of Commerce, College of Business Administration, Niagara University. His research interests include withholding effort in work teams, electronic monitoring and surveillance, ethical issues in leadership and management, and performance appraisals.
Stevens and Campion created the Teamwork Test to be used for the selection of high potential individuals into team situations. The idea of being able to test and select high team performers based on knowledge, skills, and ability (KSA) measures has great appeal. However, methods used to validate the test leave some question as to whether it is measuring capabilities related to teamwork. This study further evaluated the test by measuring the relationship between team levels of KSAs and team effectiveness.

The accumulated research evidence has shown that measures of task knowledge and task-related skills and abilities can be used to predict high levels of individual performance. The meta-analysis work of Schmidt and Hunter (1980, 1983) demonstrated significant correlations between individual levels of knowledge, skills, and abilities (KSAs) and performance evaluations. The success of KSAs in identifying important individual factors that facilitate work behavior has led to the development of selection tests specifically targeted at assessing job relevant knowledge and abilities. These tests have then been used in organizations to select individuals with high performance potential.

With the increasing popularity and use of teamwork within organizations (Bishop, Scott, & Burroughs, 2000), the development of a selection test able to assess KSAs relevant to team performance would be a distinct advantage to selection processes. Although there are many KSAs that could possibly affect teamwork, two researchers, Stevens and Campion (1994, 1999), chose and vali-
dated a number of characteristics that they suggested identify “team players.” The intent of their research was to find an appropriate domain of KSAs that could be used to select individuals whose capabilities would increase team functioning. However, Stevens and Campion evaluated the effectiveness of their teamwork selection test at the individual level. Consequently, it is difficult to know whether selecting team members based on their teamwork KSAs will actually improve team performance. The purpose of this study was to follow up on the research of Stevens and Campion and to further analyze the effectiveness of their teamwork test to predict team performance.

THEORETICAL DEVELOPMENT
OF TEAMWORK KSAs

In 1994, Stevens and Campion developed theoretical arguments and resolutions concerning a set of KSA factors that they believed would identify individuals with high levels of teamwork capabilities. Through a search of sociotechnical systems theory, organizational behavior literature, and social psychology, they inferred a set of individual-level competencies that would be effective within teamwork situations. These competencies were set within two domains. For the first domain, called Interpersonal KSAs, Stevens and Campion posited that team effectiveness would rely heavily on the ability of its individuals to successfully manage and create amicable interpersonal relations with others in the group. Hence, individual levels of Interpersonal KSAs should be strongly associated with team performance. Three factors were identified that related to team members’ abilities to handle interpersonal issues: Conflict Resolution skills (including KSAs such as recognizing types of conflict and encouraging useful conflict); Collaborative Problem Solving skills (implementing the appropriate amount of participative problem solving); and Communication skills (understanding open communication methods). In their paper, Stevens and Campion proposed that abilities on each of these factors were important to achieving good interpersonal relations within the team.
and that choosing members with high skills in conflict resolution, collaboration, and communication should create more effective teams.

In addition to interpersonal competencies, Stevens and Campion argued that a second domain of KSAs, called Self-Management KSAs, were also important to successful teamwork. Although keeping good relationships within the team should make for comfortable interactions between members, Stevens and Campion suggested that team members also needed self-management abilities to be able to direct their actions and to execute assigned tasks. They selected two Self-Management KSAs they thought critical to teamwork: Goal Setting and Performance Management (knowledge of goal setting and feedback), and Planning and Task Coordination (coordinating team member activities).

In total, Stevens and Campion identified five teamwork factors within the two domains of interpersonal and self-management skills that would be critical to creating high team performance: Conflict Resolution, Collaborative Problem Solving, Communication, Goal Setting and Performance Management, and Planning and Task Coordination. They indicated that each of these capabilities is likely to be important to producing a well-completed team project and that selecting individuals for teamwork situations who have tested high on these KSAs should result in more effective groups. A more thorough description of each of the five key factors within these two domains is presented in Table 1.

EVALUATION OF THE TEAMWORK SELECTION TEST

Based on their theoretical arguments, Stevens and Campion (1999) developed and assessed a multiple-choice paper-and-pencil test called the Teamwork Test, which was designed to measure the five team-performance factors. The test focused on teamwork knowledge and presented test takers with situational questions relating to teamwork experiences. This test was evaluated in two validation studies. The following reviews the studies and results of Stevens and Campion’s research.
In study 1, the Teamwork Test was administered to all employees of a newly opened and newly staffed pulp mill. The incremental validity of the Teamwork Test was evaluated against traditional employment aptitude tests measuring verbal, quantitative, perceptual, and mechanical reasoning abilities. The criterion validity measures consisted of supervisory ratings of individual effectiveness on what was referred to as the technical or task performance (an average of scores on technical knowledge and learning orientation) and supervisor’s ratings of the individual’s team performance (an average of scores on self-management, team contribution, and communication). Ideally, the Teamwork Test scores should be more highly correlated with team performance ratings than with task performance ratings. Also, when compared to scores from the cognitive aptitude tests, the Teamwork Test scores should be better predictors of the supervisor’s teamwork performance ratings. Instead, the results of Stevens and Campion’s first study showed

<table>
<thead>
<tr>
<th>TABLE 1: The Knowledge, Skills, and Abilities (KSAs) Measured by the Teamwork Testa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Interpersonal KSAs</strong></td>
</tr>
<tr>
<td>2. Collaborative Problem Solving KSAs. Identifying situations requiring participative group problem solving and using the proper degree of participation; and recognizing obstacles to collaborative group problem solving and implementing appropriate corrective actions.</td>
</tr>
<tr>
<td>3. Communication KSAs. Understanding effective communication networks and using decentralized networks where possible; recognizing open and supportive communication methods; maximizing the consistency between nonverbal and verbal messages; recognizing and interpreting the nonverbal messages of others; and engaging in and understanding the importance of small talk and ritual greetings.</td>
</tr>
<tr>
<td><strong>Self-management KSAs</strong></td>
</tr>
<tr>
<td>4. Goal Setting and Performance Management KSAs. Establishing specific, challenging, and accepted team goals; and monitoring, evaluating, and providing feedback on both overall team performance and individual team member performance.</td>
</tr>
<tr>
<td>5. Planning and Task Coordination KSAs. Coordinating and synchronizing activities, information, and tasks between team members, as well as aiding the team in establishing individual task and role assignments that ensure the proper balance of workload between team members.</td>
</tr>
</tbody>
</table>

a. Adapted from Stevens and Campion (1994, p. 505).
that the Teamwork Test predicted teamwork \((r = .44)\) less effectively than it predicted task work \((r = .56)\), and it predicted teamwork performance only marginally better than the aptitude tests \((p < .10)\). As well, the Teamwork Test measured technical performance at the same level as the traditional aptitude battery. On the other hand, the assessment of incremental validity for the Teamwork Test showed that there was a significant increase over the cognitive test in measuring the supervisor’s rating of the individual’s teamwork performance \((\text{incremental } R^2 = .08)\), but there was some question as to whether this increase in incremental validity was simply due to the increased reliability of measurement on the single construct of cognitive aptitude or whether the additional construct of teamwork aptitude was being measured. The results of study 1 were somewhat inconclusive as to whether the Teamwork Test provides any advantage to individual cognitive tests when predicting team capabilities.

In study 2, Stevens and Campion assessed the validity of the Teamwork Test against the performance of a sample of employees at a cardboard box plant. The purpose of this study was to provide a replication of study 1 and to evaluate teams that had been working together for a longer period of time. Again, criterion measures were obtained from supervisors, although they were modified somewhat from the first study. Three task measures (technical knowledge depth, technical knowledge breadth, and learning orientation) were taken. Also, the teamwork criterion measures were expanded so that the teamwork ratings more closely reflected the factors assessed by the Teamwork Test (Resolving Conflicts, Collaborative Behaviors, Interpersonal Communication, Goal Setting, Performance Management, and Coordinating and Planning). In addition to measuring the supervisor’s evaluation of the individual’s performance, measures of the relationship between self-ratings and the criterion measures, and between peer nominations and criterion measures, were also taken. The results of this study showed that the Teamwork Test was correlated at about the same level with both the supervisor’s task rating \((r = .25)\) and the teamwork rating \((r = .21)\) indicating that the test was equally effective at predicting both. Also, there was no significant difference between the correlations
of the Teamwork Test \((r = .21)\) and of the aptitude composite \((r = .23)\) with the supervisor’s teamwork performance evaluation. Finally, there was a nonsignificant relationship between the Teamwork scores and self teamwork ratings, but there was a positive and significant correlation between the Teamwork scores and peer teamwork nominations. Independently, the results of this study and of the first study suggest that, in general, the Teamwork Test did not predict team performance any better than nor have any added predictive validity above that of traditional cognitive tests. However, as a final evaluation of the Teamwork Test, Stevens and Campion combined the results of study 1 and study 2 and found that the Teamwork Test had significant incremental validity to predict team performance, above that of the traditional employment aptitude tests \((\text{incremental } R^2 = .03)\).

**INTERPRETING STEVENS AND CAMPION’S RESULTS**

Although the idea of being able to test for teamwork abilities is very attractive, the outcomes of the above evaluations of the Teamwork Test were mixed and leave some question as to whether the test is actually evaluating teamwork-related potential over and above any individual-level performance capabilities. In general, the Teamwork Test seemed to be equally good at measuring individual technical capabilities as it was at measuring team performance. These results suggest that the test may be measuring only individual performance factors rather than team factors. Stevens and Campion (1999) themselves posited that the correlations between the traditional cognitive test battery and the Teamwork Test \((\text{ranging from } .91 \text{ to } .99)\) could indicate that the test is simply measuring general cognitive ability. However, they also suggested that cognitive ability may be important to team performance.

At the same time, although individual cognitive capabilities are important to performance, teamwork has various other dynamics that could have an impact on productivity. Many of us will have memories of the very bright, but shy, team member whose potential
contributions went unrecognized in the team situation while the more dominant, but possibly less capable, members of the team steered the direction of the outcomes. Thus, it is not just knowledge, or in this case knowledge of correct teamwork practices, that produces high performance but also the capability to put that knowledge into play. Due to general findings of high correlations between job knowledge tests and job performance (Schmidt & Hunter, 1998), developers of paper-and-pencil selection tests might assume that an individual high on knowledge will also have high job performance. However, some job tasks and performance capabilities may not be well evaluated using paper-and-pencil knowledge tests. In the case of interpersonal or teamwork skills, individuals may be knowledgeable but not have the ability to implement their knowledge. Thus, we are still left with the question of whether the Teamwork Test is useful for predicting team performance. In addition, not just individual capabilities but also the distribution of teamwork skills throughout the group may be important to determining group performance. This issue was acknowledged by Stevens and Campion (1999) but not addressed within their studies.

One difficulty in knowing whether there is a relationship between the Teamwork Test and team performance is because all measures were taken at the individual level. Stevens and Campion (1999) themselves wondered whether their method of taking criterion measures had caused evaluators to confuse task and teamwork performance. The supervisors, assessing at the individual level, may not have been able to distinguish between individual and team relevant capabilities, and therefore this method of evaluation may not have been a good check of the Teamwork Test’s potential. Although Stevens and Campion (1999) concluded that the Teamwork Test is useful for predicting team-related performance behaviors, their assessment did not include any data on team levels of KSAs or team performance as a criterion measure. Therefore, it is difficult to truly know whether selecting people with high team KSA scores will actually improve group performance. The evidence thus far leaves open the question as to whether the test actually measures capabilities unique to teamwork or whether it just increases the amount of explained predictive variance when select-
ing for individual cognitive capabilities and individual performance. Since the purpose of the Teamwork Test is to increase team capabilities, it is important to establish whether the test is related to team outcomes. The purpose of this study, therefore, was to further evaluate the Teamwork Test by measuring the relationship between team-level KSAs and team-level measures of effectiveness.

REEVALUATING THE TEAMWORK TEST AND TEAM OUTCOMES

Although there are always measurement concerns when combining individual scores into team scores, the intent of the Teamwork Test is to identify high scoring individuals with the purpose of hiring the top scorers into teamwork situations. As with any aptitude test, the assumption is that higher test scores will lead to better performance. Thus, the consequence of selecting individuals with high aptitude levels will be a higher team-level KSA score that, in turn, should result in better team performance. The very purpose of the Teamwork Test suggests that it is most desirable to have high team levels on the KSA scores. Therefore, if the Teamwork Test is an effective method to measure and select for team performance capabilities, it would be expected that there would be a positive correlation between the team KSA score and team performance.

Hypothesis 1: There will be a significant positive relationship between the team score on the Teamwork Test and team performance.

Team levels of effectiveness can be measured in several ways. Each of the KSA factors is related to group process functions such as conflict resolution and communication skills. Each of these factors should then enhance the interactions between team members. Consequently, it would be expected that team KSA scores should be positively related to both satisfaction and performance levels in the group. Therefore, high team levels on the Teamwork Test should predict high levels of team satisfaction, as well as high levels of team performance.
Hypothesis 2: There will be a significant positive relationship between the team score on the Teamwork Test and team satisfaction.

In addition, absolute levels of the KSA in the team may not provide the whole picture of the team capabilities. Therefore, as well as the absolute levels of team KSAs, the effects of within-team variance could also have an impact on the outcomes. Although in some cases diversity in groups can lead to more creativity and better team performance, in this instance the Teamwork Test is measuring capabilities key to high levels of performance. Therefore, it should be expected that the higher the individual scores on these factors, the better the group performance. It is the level of KSAs that would seem to be critical in this instance. A group with homogeneously high scores should outperform groups that have either heterogeneous scores or homogeneously low scores. In addition, a heterogeneous group should outperform a homogeneously low scoring group because the team will have at least some members with higher team capabilities. Therefore, an interaction effect could be expected so that high KSA, high homogeneous groups would outperform heterogeneous groups, which would then outperform low KSA, high homogeneous groups.

Hypothesis 3: Highly homogeneous groups with high team KSA scores will outperform heterogeneous groups, which will outperform highly homogeneous groups with low team KSA scores.

With regard to satisfaction levels, Smith et al. (1994) suggested that group heterogeneity interacts negatively with group processes because it creates conflict and inhibits cohesion and communication. In addition, people are attracted to and have higher levels of satisfaction with others who are similar rather than different from themselves. Diversity creates feelings of distance between group members and can cause some members to withhold information or involvement (Eisenhardt & Schoonhoven, 1990; Hambrick & D’Aveni, 1992). In the case of KSA scores, high levels of variance between individuals indicate wide differences in skill capabilities that will likely decrease the smoothness of interaction processes and decrease satisfaction between team members. Thus, it is
expected that negative affect will increase as the variance between team member scores increases. As within-team KSA variance increases, team satisfaction levels should decrease.

**Hypothesis 4:** There will be a significant negative relationship between team variance on the Teamwork Test and team satisfaction.

### RESEARCH METHODOLOGY

**SAMPLE**

One hundred seventy-six undergraduate management majors participated in this research. Age of participants ranged from 19 to 47 years ($M = 21.5$). The students worked in groups ranging from three to five members. Out of a total of 44 teams, 24 teams had four members, 13 teams had five members, and 7 teams had three members. Participation was voluntary. As a consequence, several groups were dropped from the sample due to a lack of information. This left 42 groups in the final sample.

**MEASURES**

_Independent variables—KSAs._ The independent variables consisted of the team average score and the team variance score calculated from Stevens and Campion’s (1994) Teamwork Test. The test contains 35 self-report multiple-choice items evaluating the Interpersonal KSAs of Conflict Resolution, Collaborative Problem Solving, and Communication, as well as the Self-Management KSAs of Goal Setting and Performance Management, and Planning and Task Coordination. A test of the questionnaire’s reliability has produced an internal consistency measure of .80 (Stevens & Campion, 1999). The Teamwork Test was filled out by participants at the beginning of the course.

_Criterion measures._ Two criterion measures were taken in this study. The two measures consisted of group project grades and a
self-report of satisfaction with the group. The project grade consisted of a single measure, a percentage grade, given by the instructor for a paper submitted by each group at the end of the project. This mark was standardized to account for differences in grading between the two instructors teaching the sections.

The second criterion measure, a self-report satisfaction score, was collected on the day that the projects were handed in. The satisfaction score was created from two items assessing satisfaction with the team member interactions and with the group experience. The rating scale ranged from 1 (extremely disappointing) to 5 (excellent). A team average score was calculated. However, prior to combining individual satisfaction scores into a team score, the Interclass Correlation Coefficient (ICC) was evaluated (Shrout & Fleiss, 1979). The ICC measure tests how much of the construct’s total variance was due to the group-level properties of the data and tests within-group and between-group agreement. The ICC score was .52. In addition, an estimate of the convergence of satisfaction within the groups was assessed using James, Demaree, and Wolf’s (1984) within-group agreement ($r_{wg}$) measure. Values of .70 or higher are necessary to demonstrate homogeneity within the group. The $r_{wg}$ value for group measures of satisfaction was .78. These analyses provided sufficient indications that there were differences between groups and agreements within groups on levels of satisfaction. Thus, groups can be differentiated on satisfaction and it is acceptable to use an aggregate mean value for each team.

THE TASK

As a course requirement, participants were involved in an organizational simulation. In this simulation, students performed as members of an organization and solved problems related to organizational design. On the first day of the term, everyone was given a description of the organization, its departments, and its situation. Additional situational problems were added throughout the course. The students divided themselves up into one of four departments within the organization, and members of a department worked together to prepare solutions to the problems of that department.
One individual within the department volunteered to be the department head. This person was then given the leadership position within the team. At the end of the project, each group submitted for grading a report outlining their analysis and solutions to the departmental and organizational problems. The context presented to the student groups was similar to that faced by organizational project teams in that the teams were organized to solve a problem and were disbanded once the project was completed.

RESULTS

HYPOTHESES MAIN EFFECTS

The summary statistics and zero-order correlations between the team-level measures on the Teamwork Test and team outcomes can be found in Table 2. The corresponding standardized Beta scores, F, and R² statistics can be found in Table 3. Team size was used as a covariate in the analysis.

Hypotheses 1, 2, and 4 predicted main effects between the Teamwork Test and team outcomes. These hypotheses were not supported. There were no significant relationships between the team-level test score and either performance or satisfaction outcomes. There was also no significant relationship between group variances on the Teamwork Test and the satisfaction levels of the groups.

HYPOTHESES INTERACTION EFFECT

Hypothesis 3 proposed an interaction between team variance on the Teamwork Test and team performance. There was an indication of an interaction between the team scores and variances, but this finding marginally failed to achieve significance at the 5% level (p = .07). In addition, the interaction effect that did appear was unexpected. Teams with high Teamwork Test variances and high Teamwork scores outperformed all others (M = .48), whereas teams with low Teamwork Test variances and high test scores were the
<table>
<thead>
<tr>
<th>Scales</th>
<th>M</th>
<th>SD</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Teamwork test</td>
<td>22.10</td>
<td>2.76</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>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Teamwork test variance</td>
<td>18.17</td>
<td>16.47</td>
<td>-.37**</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>
<td></td>
</tr>
<tr>
<td>3. Conflict resolution (CR)</td>
<td>2.89</td>
<td>.51</td>
<td>.40***-.48***</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>
<td></td>
</tr>
<tr>
<td>4. CR variance</td>
<td>.68</td>
<td>.66</td>
<td>-.23</td>
<td>.16</td>
<td>-.36***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Problem solving (PS)</td>
<td>5.92</td>
<td>6.70</td>
<td>.15</td>
<td>.06</td>
<td>.06</td>
<td>.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. PS variance</td>
<td>3.19</td>
<td>2.73</td>
<td>-.07</td>
<td>.50***-.33**</td>
<td>-.02</td>
<td>-.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Communication skill (CS)</td>
<td>8.46</td>
<td>9.31</td>
<td>.11</td>
<td>.12</td>
<td>-.08</td>
<td>-.12</td>
<td>-.03</td>
<td>-.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. CS variance</td>
<td>3.58</td>
<td>3.40</td>
<td>-.29</td>
<td>.50***-.30*</td>
<td>-.20</td>
<td>-.01</td>
<td>.24</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Performance management (PM)</td>
<td>3.31</td>
<td>.71</td>
<td>.49***-.16</td>
<td>.20</td>
<td>-.24</td>
<td>.05</td>
<td>-.13</td>
<td>.20</td>
<td>-.54***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. PM variance</td>
<td>1.63</td>
<td>1.78</td>
<td>.40***</td>
<td>.56***-.35**</td>
<td>-.05</td>
<td>-.12</td>
<td>.59***-.16</td>
<td>.20</td>
<td>-.34**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Planning and coordination (PC)</td>
<td>3.92</td>
<td>.76</td>
<td>.65***</td>
<td>-.06</td>
<td>.14</td>
<td>.15</td>
<td>.20</td>
<td>-.08</td>
<td>.18</td>
<td>-.31**</td>
<td>.35**</td>
<td>-.25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. PC variance</td>
<td>.96</td>
<td>.73</td>
<td>-.41***</td>
<td>.46***-.33**</td>
<td>.30</td>
<td>-.16</td>
<td>.27</td>
<td>-.05</td>
<td>.36**</td>
<td>-.34</td>
<td>.40***</td>
<td>-.31**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Project grade (standardized)</td>
<td>.01</td>
<td>.94</td>
<td>.20</td>
<td>.23</td>
<td>-.26</td>
<td>-.21</td>
<td>.20</td>
<td>.13</td>
<td>.10</td>
<td>.12</td>
<td>.12</td>
<td>.06</td>
<td>.30**</td>
<td>.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Team satisfaction</td>
<td>3.71</td>
<td>.60</td>
<td>.10</td>
<td>-.05</td>
<td>.43***</td>
<td>-.20</td>
<td>.09</td>
<td>.23</td>
<td>.06</td>
<td>.03</td>
<td>-.26*</td>
<td>.20</td>
<td>.07</td>
<td>-.02</td>
<td>-.16</td>
<td></td>
</tr>
<tr>
<td>15. Group size</td>
<td>4.12</td>
<td>.80</td>
<td>-.01</td>
<td>-.04</td>
<td>.04</td>
<td>-.08</td>
<td>.15</td>
<td>-.11</td>
<td>.01</td>
<td>.11</td>
<td>-.08</td>
<td>-.07</td>
<td>-.11</td>
<td>.12</td>
<td>.13</td>
<td>.12</td>
</tr>
</tbody>
</table>

NOTE: n = 42 groups.

* p < .10. ** p < .05. *** p < .01.
worst performing teams ($M = -.06$). This result suggests that the
distribution of the KSA capabilities is more important than having a
high overall KSA capability within the team.

**FURTHER EXPLORATION OF THE TEAMWORK TEST**

In developing the theoretical rational for the Teamwork Test,
Stevens and Campion (1994) created propositions for each of the
five component teamwork factors of Conflict Resolution, Collabora-
tive Problem Solving, Communication, Goal Setting and Perfor-

<table>
<thead>
<tr>
<th>Variable</th>
<th>Performance</th>
<th></th>
<th></th>
<th>Satisfaction</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\beta$</td>
<td>$F$</td>
<td>$R^2$</td>
<td>$\beta$</td>
<td>$F$</td>
<td>$R^2$</td>
</tr>
<tr>
<td>Group size</td>
<td>.15</td>
<td>.74</td>
<td>.02</td>
<td>.12</td>
<td>.53</td>
<td>.01</td>
</tr>
<tr>
<td>Teamwork test</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team score</td>
<td>.20</td>
<td>1.18</td>
<td>.02</td>
<td>.10</td>
<td>.46</td>
<td>.02</td>
</tr>
<tr>
<td>Team variance</td>
<td>.23</td>
<td>1.53</td>
<td>.07</td>
<td>-.00</td>
<td>.30</td>
<td>.02</td>
</tr>
<tr>
<td>Team score $\times$ team variance</td>
<td>.36</td>
<td>2.55*</td>
<td>.17</td>
<td>-.04</td>
<td>.30</td>
<td>.02</td>
</tr>
<tr>
<td>Conflict resolution</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team score</td>
<td>-.27</td>
<td>1.93</td>
<td>.09</td>
<td>.43</td>
<td>5.28***</td>
<td>.22</td>
</tr>
<tr>
<td>Team variance</td>
<td>.22</td>
<td>1.43</td>
<td>.05</td>
<td>.06</td>
<td>.99</td>
<td>.05</td>
</tr>
<tr>
<td>Team score $\times$ team variance</td>
<td>.18</td>
<td>1.40</td>
<td>.11</td>
<td>.38</td>
<td>2.69**</td>
<td>.18</td>
</tr>
<tr>
<td>Collaborative problem solving</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team score</td>
<td>.18</td>
<td>1.04</td>
<td>.02</td>
<td>.09</td>
<td>.36</td>
<td>.02</td>
</tr>
<tr>
<td>Team variance</td>
<td>.15</td>
<td>.80</td>
<td>.04</td>
<td>.15</td>
<td>.65</td>
<td>.03</td>
</tr>
<tr>
<td>Team score $\times$ team variance</td>
<td>.25</td>
<td>1.75</td>
<td>.08</td>
<td>.17</td>
<td>.81</td>
<td>.04</td>
</tr>
<tr>
<td>Communication</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team score</td>
<td>.10</td>
<td>.54</td>
<td>.03</td>
<td>.06</td>
<td>.34</td>
<td>.00</td>
</tr>
<tr>
<td>Team variance</td>
<td>.11</td>
<td>.60</td>
<td>.03</td>
<td>.01</td>
<td>.26</td>
<td>.01</td>
</tr>
<tr>
<td>Team score $\times$ team variance</td>
<td>.12</td>
<td>.48</td>
<td>.04</td>
<td>.07</td>
<td>.34</td>
<td>.02</td>
</tr>
<tr>
<td>Performance management</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team score</td>
<td>.17</td>
<td>.70</td>
<td>.03</td>
<td>-.21</td>
<td>1.64</td>
<td>.08</td>
</tr>
<tr>
<td>Team variance</td>
<td>.13</td>
<td>.46</td>
<td>.02</td>
<td>.14</td>
<td>1.18</td>
<td>.06</td>
</tr>
<tr>
<td>Team score $\times$ team variance</td>
<td>.06</td>
<td>.66</td>
<td>.03</td>
<td>.17</td>
<td>.83</td>
<td>.04</td>
</tr>
<tr>
<td>Planning and task coordination</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Team score</td>
<td>.32</td>
<td>2.69*</td>
<td>.12</td>
<td>.09</td>
<td>.44</td>
<td>.02</td>
</tr>
<tr>
<td>Team variance</td>
<td>.17</td>
<td>.95</td>
<td>.05</td>
<td>.00</td>
<td>.28</td>
<td>.01</td>
</tr>
<tr>
<td>Team score $\times$ team variance</td>
<td>.53</td>
<td>3.12**</td>
<td>.20</td>
<td>-.06</td>
<td>.32</td>
<td>.02</td>
</tr>
</tbody>
</table>

*p < .10. ** p < .05. *** p < .01.
mance Management, and Planning and Task Coordination. They proposed that each of these factors was positively related to team effectiveness. Therefore, the effects of scores on each of these teamwork factors were separately investigated and the results can be seen in Tables 2 and 3.

An examination of the regression analyses for each of the team factors showed two significant effects. Team Conflict Resolution had a significant positive relationship to team satisfaction \( (p = .005; R^2 = .22) \). As the team level of Conflict Resolution knowledge increased, so did the satisfaction level of the team. In addition, there was a significant interaction effect for the team score and variance levels of the Planning and Task Coordination KSA \( (p = .03; R^2 = .20) \). The highest performing groups were homogeneous, with high team knowledge of planning and coordination \( (M = .20) \), whereas the worst performers were homogeneous teams with low scores on planning and coordinating \( (M = -.59) \). This result was consistent with what had been hypothesized (Hypothesis 3) as the expected interaction for the total Teamwork Test.

**DISCUSSION**

Selection is one of the cornerstones of human resource practices. With organizations more than ever before using teamwork to gain competitive advantage, the capability to select good team players has become increasingly important. Stevens and Campion (1999) developed a paper-and-pencil knowledge test to evaluate KSA relevant to teamwork situations and for the purpose of personnel selection. When the validity of the Teamwork Test was analyzed in earlier studies (Stevens & Campion, 1999), the potential of the test to predict performance appraisals of individual teamwork was supported, but the support was not strong. The current study reevaluated the predictive validity of the test by analyzing variables at the team level. This study found a nonsignificant relationship between the team scores on Stevens and Campion’s Teamwork Test and group effectiveness. Thus, the results did not give support to the Teamwork Test as a tool useful for predicting team performance.
The combined findings of these studies puts in doubt the validity of the Teamwork Test for selecting people into team situations.

Stevens and Campion (1999) have suggested that the weak results of their validation studies may have been because the test is measuring general cognitive abilities of the individual participants rather than any new dimension of teamwork. This suggestion provides some explanation as to why the test scores obtained in their study were predictive of ratings on both technical and team performance and why the cognitive tests and Teamwork Test were so highly correlated (.91 - .99). The Teamwork Test may actually be measuring individual rather than team potential. Stevens and Campion (1999) also suggested that their criterion measure (for example, supervisor ratings of an individual’s team performance) may have caused confusion such that raters were actually rating the more traditional individual performance rather than team capabilities. By measuring outcomes at the team level, this study attempted to overcome some of the above problems created by individual performance ratings and to provide a better evaluation of the test’s usefulness to predict team outcomes. Thus, the differences in the methodologies of the two studies may account for some of the disparity in results between the Stevens and Campion study and this research. Measures used here were taken at the group level. Inasmuch as the tool is intended to predict and improve team performance, testing at the team level may have provided a better evaluation of the tool’s capability to predict team performance and overcome some of the concerns expressed by Stevens and Campion with regard to their methodology. This is not to suggest that the aggregated measure creates a group identity, but rather the method used in this study allowed testing of Stevens and Campion’s framework for team member selection. Their framework suggested that high individual scores on the Teamwork Test indicated high team capabilities. Consequently, when measuring at the team level, those teams composed of high scoring individuals should outperform teams composed of low scoring individuals. In other words, there should be a positive linear relationship between the team level test score and team performance. This hypothesis was not supported by the results of the study.
Whereas the main effects tests for a relationship between test scores and outcome measures did not yield significant results, the interaction effects of the team test score and team variance score did produce a weak effect. High test score, high variance teams outperformed all others, whereas high test score, low variance teams performed the worst. This result also gives some support to the contention that the Teamwork Test is measuring individual rather than team performance characteristics. Team variance on KSAs appeared to be more important to performance than having a high overall KSA capability. If the Teamwork Test is predictive of team performance, it would be expected that high scoring, homogeneous groups would be the best performers. However, if the test is measuring non-team-related, individual characteristics, then diversity rather than homogeneity would be expected to increase team performance. This is consistent with previous research that has found that diversity in individual differences is positively related to team outcomes when teams are dealing with complex tasks (Bowers, Pharmer, & Salas, 2000). Diversity brings value to a group by providing alternate viewpoints, ideas, and methods, which ultimately improve performance (Cox, Lobel, & McLeod, 1991). Thus, effects of variances on the test measures may be due to bringing different methods and ideas into the group rather than due to any direct relationship between levels of KSAs and teamwork performance.

Nonetheless, when the independent KSA factors were examined, there were several noteworthy results. As might be expected, high team levels of Conflict Resolution did appear to smooth the interaction processes between team members. The results showed a positive association between conflict resolution capabilities and levels of team satisfaction. It could be stated that conflict management capabilities may enhance group processes by focusing team members on creating internal harmony.

A second significant result that appeared when examining the KSA factors was an interaction effect between team homogeneity and team score for Planning and Task Coordination on team performance. Planning and Task Coordination measured the team members’ understanding of the need to coordinate and synchronize team
member activities, as well as the need to balance member workloads. The highest performing groups were homogeneous, with high team knowledge of planning and coordination, whereas the worst performers were homogeneous teams with low scores on planning and coordinating. The results of this study indicated that as the team member capabilities for coordination increased, so did team performance.

LIMITATIONS AND FUTURE RESEARCH

The limitations of this study provide an opportunity for future research. The validation work by Stevens and Campion (1999) was done in organizations. Unfortunately, although teamwork has been implemented into many organizations, few organizations have adopted evaluation systems at the team level. The majority of evaluation systems still assess individual performance. Consequently, it is difficult to obtain team performance measures. Using student groups provided the opportunity to gain team-level performance information. The design of this study also overcame some of the weaknesses reported in the Stevens and Campion (1999) study. For example, these teams provided an intensive team situation in which all were engaged in a similar cognitively complex task. Hence, using student groups rather than organizational groups allowed for some situational controls that Stevens and Campion could not achieve in organizations. In their recent review of methodological trends, Scandura and Williams (2000) noted the move away from studies allowing for control of the environment in favor of studies of high contextual realism. They cautioned that there are trade-offs in all methods with regard to precision of measurement, realism, and generalizability of findings, and recommended that the use of a variety of methods to examine a topic should result in more robust and generalizable findings. This study, in addition to the study of Stevens and Campion (1999), helps to achieve that objective. Last, it would be expected that the factors measured by the Teamwork Test (e.g., task coordination, communication, and task management) are relevant to all group work. The test, therefore, should predict team performance in both student and organizational groups.
Although the lack of relationship between the overall test scores and student performance creates some question as to the generalizability of these measures to work group performance, this study adds to our current knowledge of the Teamwork Test and leaves the door open to a future study that will expand the target sample into organizations using team-level evaluations and more permanent teams.

Another issue that may be addressed in future research is how variances in work interdependence affect the relationship between test scores and performance outcomes. Because a single type of task was used, no specific measures of work interdependence were gathered in this study. However, information collected during the study on the number of meetings held by different groups suggests that teams varied on the degree to which members worked together or independently. A post hoc examination of the relationship between the team test score, the number of team meetings, and team outcomes showed a significant correlation between the number of meetings and team performance \( (r = .32, p < .05) \) but no relationship between the Teamwork Test score and the number of team meetings \( (r = -.06, p < .75) \). Whereas these results do indicate that the more interdependently the team worked the better the performance, it also showed no indication that the aggregated Teamwork Test score was related to the degree to which members decided to work interdependently. This post hoc analysis was a rough measure of the level of interdependence used by the team members to accomplish the task. However, an area of further investigation would be to explore whether the Teamwork Test will differentially predict teamwork performance depending on the level of interdependence required by the task.

**CONCLUSION**

This study addressed some of the issues put forward by Stevens and Campion (1999) as directions for future research. The tasks carried out by teams were cognitively complex, the effects of team heterogeneity were assessed, the test was applied in a team applica-
tion quite different from that studied by Stevens and Campion, and
the criteria were examined at a team rather than individual level. 
The results of this investigation of the Teamwork Test generally
found that high team-level KSA scores did not produce better group
performance. It is therefore suggested that the Teamwork Test may
be measuring some individual capabilities but that these character-
istics are better predictors of individual rather than team-level per-
formance. To further evaluate the Teamwork Test, it would be
appropriate to administer it to team members in organizations that
perform appraisals at the team level or that use team-level measures
of productivity.

Finally, the findings of this study, combined with those of
Stevens and Campion’s results, leave doubt as to whether the Team-
work Test is truly predictive of team performance characteristics
and whether using the Teamwork Test as a selection tool would pro-
duce better performing teams. In theory, Stevens and Campion
(1994) proposed that a number of KSA factors were important to
the performance of individuals engaged in teamwork. The intent of
this study was not to suggest that these teamwork skills are unim-
portant to team performance. In fact, Stevens and Campion’s
(1994) review of the literature suggests that conflict resolution, col-
laborative problem solving, communication, goal setting and per-
formance management, and planning and task coordination are
competencies that are influential to team outcomes. However, the
selection technique developed by Stevens and Campion (1999) to
assess these teamwork capabilities may not have been appropriate.
A number of team skills are interpersonal and interaction skills, and
although a paper-and-pencil knowledge test such as the Teamwork
Test is a very efficient method to assess these capabilities, it may
not be any more effective in assessing teamwork capabilities than
other general cognitive tests. Knowledge, by itself, does not neces-
sarily indicate performance proficiencies or the mastery of com-
plex interpersonal skills. Anderson’s (1983) ACT* theory of cogni-
tion proposes three stages through which skills acquisition must
proceed: (a) a cognitive stage, in which a description of the proce-
dure is learned (declarative knowledge); (b) an associative stage, in
which the facts are compiled and integrated with a method for per-
forming the skills (procedural knowledge); and (c) an autonomous stage, in which the skill becomes more fluent and automatic. The third stage is crucial for competent performance (May & Kahnweiler, 2000). Although a knowledge test may evaluate stages one and two of the steps toward the acquisition of interpersonal skills, it is less able to evaluate the final stage of behavioral mastery. Techniques other than knowledge tests might better evaluate teamwork behaviors. For example, at assessment centers, tests such as role plays and leaderless group discussions are used to determine interpersonal skills. These methods do assess behaviors and may be more effective and appropriate approaches for assessing team performance capabilities and for selecting individuals into group work situations.

REFERENCES


Diane L. Miller is an assistant professor in the Faculty of Management at The University of Lethbridge. She received her Ph.D. in organizational behavior from the University of Toronto. Her current research interests include team processes and performance, the impact of diversity within groups, and feedback-seeking behavior.
INDEX

TO

SMALL GROUP RESEARCH

Volume 32

Number 1 (February 2001) 1-112
Number 2 (April 2001) 113-256
Number 3 (June 2001) 257-376
Number 4 (August 2001) 377-500
Number 5 (October 2001) 501-668
Number 6 (December 2001) 669-772

Authors:

ALISON, LAURENCE J., see Porter, L. E.
BACALLAO, MARTICA L., see Smokowski, P. R.
BEAUCHAMP, MARK R., and STEVEN R. BRAY, “Role Ambiguity and Role Conflict Within Interdependent Teams,” 133.
BORDIA, PRASHANT, see Chang, A.
BRAY, STEVEN R., see Beauchamp, M. R.
BROWN, TRAVOR C., see Taggar, S.
CARRON, ALBERT V., see Colman, M. M.
CARRON, ALBERT V., see Eys, M. A.
CARRON, ALBERT V., see Gammage, K. L.
CARRON, ALBERT V., see Loughead, T. M.
PHILIPS, JENNIFER L., see Devine, D. J.
PINSONNEAULT, ALAIN, see Barki, H.
PIROLA-MERLO, ANDREW, see Bain, P. G.
PRIOLA, VINCENZA, see Armstrong, S. J.
ROSE, SHELDON D.; see Smokowski, P. R.
SUE-CHAN, CHRISTINA, see Sargent, L. D.
TAGGAR, SIMON, and TRAVOR C. BROWN, “Problem-Solving Team Behaviors: Development and Validation of BOS and a Hierarchical Factor Structure,” 698.
TOOTELL, GEOFFREY, see De Kelaita, R.

Articles:
“Does Diversity Affect Group Efficacy? The Intervening Role of Cohesion and Task Interdependence,” Sargent and Sue-Chan, 426.
“Generating Agreement in Computer-Mediated Groups,” Whitworth et al., 624.
“Improving Face-to-Face Brainstorming through Modeling and Facilitation,” Kramer et al., 533.
“The Innovation Imperative: The Relationships Between Team Climate, Innovation, and Performance in Research and Development Teams,” Bain et al., 55.
“Investigating the Mediation of Leadership, Class Cohesion, and Adherence in an Exercise Setting,” Loughead et al., 558.


“A Multidimensional Approach to the Group Cohesion–Group Performance Relationship,” Chang and Bordia, 379.


“Problem-Solving Team Behaviors: Development and Validation of a Hierarchical Factor Structure,” Taggar and Brown, 698.


“Reexamining Teamwork KSAs and Team Performance,” Miller, 745.


“Risky Decision Making in Mixed-Gender Teams: Whose Risk Tolerance Matters?,” Karakowsky and Elangovan, 94.

“Role Ambiguity and Role Conflict Within Interdependent Teams,” Beauchamp and Bray, 133.


“Self-Construals, Communication Processes, and Group Outcomes in Homogeneous and Heterogeneous Groups,” Oetzel, 19.


“Team Cohesion and Individual Productivity: The Influence of the Norm for Productivity and the Identifiability of Individual Effort,” Gammage et al., 3.


**Book Reviews:**