Journal of Personality and Social Psychology
2006, Vol. 90, No. 4, 644–651

0022-3514/06/$12.00 DOI: 10.1037/0022-3514.90.4.644

Groups Perform Better Than the Best Individuals on Letters-to-Numbers Problems: Effects of Group Size

Patrick R. Laughlin, Erin C. Hatch, Jonathan S. Silver, and Lee Boh
University of Illinois at Urbana–Champaign

Individuals and groups of 2, 3, 4, or 5 people solved 2 letters-to-numbers problems that required participants, on each trial, to identify the coding of 10 letters to 10 numbers by proposing an equation in letters, receiving the answer in letters, proposing a hypothesis, and receiving feedback on the correctness of the hypothesis. Groups of 3, 4, and 5 people proposed more complex equations and had fewer trials to solution than the best of an equivalent number of individuals. Groups of 3, 4, and 5 people had fewer trials to solution than 2-person groups but did not differ from each other. These results suggest that 3-person groups are necessary and sufficient to perform better than the best individuals on highly intellective problems.

Keywords: group problem solving, group size, intellective tasks

Cooperative groups perform better than independent individuals on a wide range of problems (for representative reviews, see Hastie, 1986; Hill, 1982; Kerr & Tindale, 2004; Levine & Moreland, 1998). This research has traditionally compared an equal number of groups and individuals (e.g., 20 four-person groups with 20 individuals), thus comparing groups and the average individual. A more stringent test of group versus individual performance is a comparison of n groups of size m with an equivalent number of $n \times m$ individuals (e.g., 20 groups of size four with 80 individuals). This allows comparison of groups with the best, second best, and so forth through the mth best of an equivalent number of individuals rather than the usual comparison of groups and the average individual. Extending traditional research to this more stringent comparison, two recent experiments reported that four-person groups (Laughlin, Bonner, & Miner, 2002) and three-person groups (Laughlin, Zander, Knievel, & Tan, 2003) performed better than the best of an equivalent number of individuals on a highly intellective task, letters-to-numbers problems.

Comparisons of the performance of cooperative groups of a given size and individuals are a special case of the larger issue of the relationship between group size and performance. The current experiment addressed this larger issue by a comparison of groups of size two, three, four, and five people and the best of an equivalent number of individuals on letters-to-numbers problems. In this article, we first explain letters-to numbers problems, consider possible strategies, and review the experiments of Laughlin et al. (2002, 2003).We then review the surprisingly small amount of previous research on the effects of group size in problem solving. From these considerations, we predicted (a) better performance for groups of each of size two, three, four, and five than an equivalent number of individuals and (b) major improvement in performance from group size two to three, with decreasing improvement from group sizes three to four to five.

Letters-to-Numbers Problems

Previous to the experiment, we randomly assigned each of the 10 letters *A*, *B*, *C*, *D*, *E*, *F*, *G*, *H*, *I*, and *J* (without replacement) to 1 of the 10 numbers 0, 1, 2, 3, 4, 5, 6, 7, 8, and 9. The objective for the problem solvers was to identify this mapping of the 10 letters to the 10 numbers in as few trials as possible. On each trial, the problem solvers proposed an equation in letters (e.g., $A + D =$?). The experimenter then gave the answer to the equation in letters (e.g., $A + D =$ B). The problem solvers then proposed one specific mapping of a letter to a number (e.g., $A = 3$) and were told whether the hypothesis was true or false (e.g., true, *A* equals *3*; or false, *A* does not equal *3*). The problem solvers then filled out the coding of the 10 letters to the 10 numbers on their answer sheets. The full correct coding solved the problem, whereas an incorrect coding required another trial. A maximum of 10 trials was allowed. The Appendix gives the instructions with four illustrative trials. How might these problems be solved? We first consider a simple but inefficient two-letter substitution strategy. We then consider more efficient multiletter strategies.

Two-Letter Substitution Strategy

Table 1 illustrates a two-letter substitution strategy. Once any one of the letters is identified by logic (e.g., the two-letter answer to the equation $A + J = ED$ means that the first letter, *E*, must be 1) or experimenter feedback on the hypothesis (e.g., true, *E* is 1), the letter may be used in a series of two-letter equations to identify the other letters. Nine of the letters have been identified by Trial 9, and the remaining letter follows by exclusion. Hence, the problem

Patrick R. Laughlin, Erin C. Hatch, Jonathan S. Silver, and Lee Boh, Department of Psychology, University of Illinois at Urbana–Champaign.

Erin C. Hatch is now at The Shepherd's Table, Silver Spring, Maryland. Jonathan S. Silver is now at the College of Education, University of Iowa. Lee Boh is now at the 3rd Divisional Air Defence Artillery Battalion, Republic of Singapore Airforce, Singapore.

Correspondence concerning this article should be addressed to Patrick R. Laughlin, Department of Psychology, University of Illinois at Urbana–Champaign, 603 East Daniel Street, Champaign, IL 61820. E-mail: plaughli@uiuc.edu

Table 1
Two-Letter Substitution Strategy

Trial	Equation	Hypothesis	Feedback
1	$A + J = ED$	$E = 1$	True
2	$E + E = D$	$D = 2$	True
3	$E + D = A$	$A = 3$	True
4	$E + A = G$	$G = 4$	True
5	$E + G = B$	$B = 5$	True
6	$E + B = F$	$F = 6$	True
7	$E + F = H$	$H = 7$	True
8	$E + H = C$	$C = 8$	True
9	$E + C = J$	$J = 9$	True

Note. $A = 3$; $B = 5$; $C = 8$; $D = 2$; $E = 1$; $F = 6$; $G = 4$; $H = 7$; $I = 0$; $J = 9$.

is solved in nine trials. For simplicity, we have illustrated a two-letter substitution strategy with the identified letter coded as 1, but a comparable strategy may be used once any of the letters is identified. For convenience of exposition, we subsequently refer to the equation on Trial 1 of a given illustration as Equation 1, the equation on Trial 2 as Equation 2, and so on.

Although a two-letter substitution strategy will generally solve the problem in the allowed maximum of 10 trials, it is clearly suboptimal. The hypotheses on each trial redundantly test what is known, and there is no substitution of known letters in previous equations to identify further letters (e.g., after the letter *E* has been identified as 1 on Equation 1, the letter *D* has been identified as 2 on Equation 2, and the letter *A* has been identified as 3 on Equation 3, substitution of these three letters in Equation 1 identifies the letter *J* as 9).

Multiletter Strategies

Table 2 illustrates a multiletter strategy. Rather than simply adding the known *E* and *D* to identify *A* as 3, Equation 3 adds combinations of *E* and *D* to identify *A* as 3, *G* as 4, and *B* as 5. Equation 4 then uses combinations of the known letters *D* and *A* to identify *F* as 6, *H* as 7, *C* as 8, and *J* as 9. The remaining letter *I* is known by exclusion. The basic insight is that multiletter equations identify more than one letter and are therefore more informative than two-letter equations.

Table 3 illustrates a sophisticated multiletter strategy. The sum of the integers from 1 to *N* is given by the formula $N(N + 1)/2$. Because the letter coded as 0 (here *I*) has no effect, the answer to Equation 1 that adds all 10 letters is known in advance to be $9(10)/2 = 45$, identifying *G* as 4 and *B* as 5. The letter *G* is then used in the multiletter Equation 2 to identify *E* as 1, *A* as 3, and *D* as 2. The letters *E*, *D*, *A*, *G*, and *B* are then used in the multiletter Equation 3 to identify *F* as 6, *H* as 7, *C* as 8, and *J* as 9. The remaining letter, *I*, is known by exclusion, and the problem is solved in three trials. In summary, the illustrative strategies of Tables 1, 2, and 3 demonstrate the basic principle that as the number of letters in an equation increases, the equation will identify progressively more letters, and hence the problem will be solved in fewer trials.

Groups Versus the Best Individuals on Letters-to-Numbers Problems

Laughlin et al. (2002) compared 82 four-person groups and an equivalent number of $82 \times 4 = 328$ individuals on one letters-to-numbers problem. The groups proposed more complex equations (letters per equation) and had fewer trials to solution than each of the best, second-best, third-best, and fourth-best individuals. In a subsequent experiment, Laughlin et al. (2003) compared 100 three-person groups and an equivalent number of 300 individuals on two successive problems under five instruction conditions: (a) standard or unconstrained, (b) letter coded as 1 known at outset, (c) letter coded as 9 known at outset, (d) use at least three letters on all equations, and (e) use at least four letters on all equations. The groups proposed more complex equations and had fewer trials to solution than each of the best, second-best, and third-best individuals. Performance for both groups and individuals was best with the letter coded as 9 known at the outset and instructions to use at least four letters on each equation, and the other three instruction conditions did not differ significantly from each other. Because with the random assignment to group or individual conditions, the best of four group members would be equivalent to the best of four individuals, these results also indicate that the groups performed better than their best member would have performed alone.

However, although the groups used more complex equations and solved in fewer trials than the best individuals, neither experiment reported the direct relation between letters per equation and trials to solution. Accordingly, we regressed trials to solution on letters per equation in each of these previous experiments. In Laughlin et al. (2002), there was an R of .41 ($p < .001$) for the single problem. In Laughlin et al. (2003), there was an R of .39 ($p < .001$) for Problem 1 and an R of .44 ($p < .001$) for Problem 2. This indicates that the use of more complex equations led to better performance in both experiments.

The current experiment compared groups and an equivalent number of individuals at each of group size two, three, four, and five on two successive letters-to-numbers problems. From the results of Laughlin et al. (2002, 2003) and the previous considerations of effective strategies on letters-to-numbers problems, we predicted that the groups would have fewer trials to solution and propose more complex equations (letters per equation) than the best of an equivalent number of individuals at each of group size two, three, four, and five.

Group Size and Problem-Solving Performance

Does problem-solving performance (a) improve linearly with increasing group size; (b) improve up to a point and then level off;

Table 2
Combined Two-Letter and Multiletter Strategy

Trial	Equation	Hypothesis	Feedback
1	$A + J = ED$	$B = 0$	False
2	$E + E = D$	$C = 0$	False
3	$EEE + EED + EDD = AGB$	$F = 0$	False
4	$DDDA + DDAA + DAAA = FHCJ$	$I = 0$	True

Note. $A = 3$; $B = 5$; $C = 8$; $D = 2$; $E = 1$; $F = 6$; $G = 4$; $H = 7$; $I = 0$; $J = 9$.

Table 3
Sophisticated Multiletter Strategy

Trial	Equation	Hypothesis	Feedback
1	$A + B + C + D + E + F + G + H + I + J = GB$	$A = 0$	False
2	$GG + GG + GG = EAD$	$C = 0$	False
3	$EDAG + BBBB = FHCJ$	$I = 0$	True

Note. $A = 3$; $B = 5$; $C = 8$; $D = 2$; $E = 1$; $F = 6$; $G = 4$; $H = 7$; $I = 0$; $J = 9$.

or (c) improve up to a point, level off, and then decrease? Indeed, is there any systematic relation between group size and problem-solving performance? Despite the theoretical and practical importance of the relation between group size and performance, there is a surprisingly small amount of previous research.

Thorndike (1937) had students in intact classes from five universities respond as individuals and then as four-person, five-person, or six-person groups to factual world knowledge items (e.g., geography, economics, politics) and items with correct answers defined by the judgment of experts (e.g., the better of two poems, the better of two paintings). Four-person groups were correct on 59% of the items, five-person groups on 61%, and six-person groups on 63%. Taylor and Faust (1952) had individuals, two-person groups, and four-person groups play the parlor game Twenty Questions under the standard procedures (starting from a known category of animal, vegetable, or mineral and asking up to a maximum of 20 yes–no questions to identify the object). They played five games on each of 3 successive days. Four-person groups solved more of the 15 problems in the allowed 20 questions than two-person groups, but there was a nonsignificant difference for the number of questions (e.g., 17) on the problems that were correctly solved. Lorge and Solomon (1959, 1960) compared various group sizes from two to seven in intact classes in different years on the Tartaglia (husbands and wives) river-crossing problem. There was no consistent relationship between percentage of solvers and group size. For example, the solution rate was 15% for three-person groups and 13% for six-person groups in one class and 66% for four-person groups and 46% for seven-person groups in another class. Similarly, Thomas and Fink (1961) found no significant differences among groups of size two, three, four, and five on the Maier and Solem (1952) horse-trading problem.

Laughlin, Kerr, Davis, Halff, and Marciniak (1975) first gave individual college students the 115 vocabulary items of the Terman (1956) Concept Mastery Test. After dichotomizing the individual scores at the median into high and low halves, they randomly assigned participants within each half to retake the same items as individuals or in cooperative groups of size 2, 3, 4, or 5. There was significant linear improvement with increasing group size for the high-ability groups but no effect of group size for the low-ability groups. Bray, Kerr, and Atkin (1978) compared male and female groups of size 2, 3, 6, and 10 on "gold dust" (modified Luchins, 1942, water jar) problems of low, medium, or high difficulty. For problems of low difficulty, there was no difference among group sizes, probably because of a ceiling effect. For problems of moderate difficulty, male groups of size 10 had more correct answers than groups of sizes 3 and 6, but female groups did not differ significantly from each other. For problems of high difficulty, there was no effect of group size, probably because of a floor effect.

In summary, previous research suggests that performance improves with increasing group size for problems of moderate difficulty that require understanding of verbal, quantitative, or logical conceptual systems, but performance has not been shown to improve with increasing group size for so-called eureka problems. Letters-to-numbers problems require knowledge of arithmetic, algebra, and logic as well as information processing over a series of trials. The previous studies of Laughlin et al. (2002, 2003) indicate that letters-to-numbers problems are challenging but not excessively difficult for the participants. Thus, previous research suggests linear improvement with increasing group size in the present experiment.

However, we have proposed that the crucial aspect of the superior performance of groups over individuals is the use of more complex, multiletter strategies rather than the simple, obvious, but less effective two-letter substitution strategy. The 328 individuals in Laughlin et al. (2002) had a probability of .71 of using two-letter equations on all trials for their single problem. The 180 individuals over the three instruction conditions—(a) unconstrained, (b) letter coded as *1* given at outset, and (c) letter coded as *9* given at outset—of Laughlin et al. (2003) had a probability of .67 of using two-letter equations on all trials on their first problem and .65 on their second problem (the participants in the other two conditions were instructed to use at least three or four letters on all equations). This gives an overall individual probability of .68 of using two-letter equations on all trials. If we assume that multiletter equations are demonstrably preferable to two-letter equations if proposed by at least one group member, the probability of the groups using a multiletter strategy is $1 - .68^N$ (where N = group size). This predicts probabilities of .54, .69, .79, and .85 that groups of sizes two, three, four, and five will use a multiletter strategy. From these considerations, we predicted a major improvement in trials to solution from two-person to three-person groups but progressively decreasing improvement from three-person to four-person to five-person groups.

Method

Participants and Design

The participants were 760 students at the University of Illinois at Urbana–Champaign who received course credit for participation. Two hundred participants were randomly assigned to solve two successive letters-to-numbers problems as individuals, 80 as 40 two-person groups, 120 as 40 three-person groups, 160 as 40 four-person groups, and 200 as 40 five-person groups. There were 20 random codings of the 10 letters to the 10 numbers. In Replications 1–10 and 21–30, the first 10 of the random codings were used for Problem 1 and the second 10 were used for Problem 2; in Replications 11–20 and 31–40, the second 10 codings were used for Problem 1 and the first 10 were used for Problem 2.

Instructions and Procedure

The instructions are given in the Appendix. The experimenter gave these instructions orally with the four illustrative trials on a blackboard, answering any questions on the procedure. In the group conditions, the members discussed to consensus on each proposed equation and hypothesis, and each member wrote the proposed group equation in letters, answer in letters, group hypothesis, and feedback on the hypothesis on his or her response sheets on each trial. Individuals followed the same procedure without discussion. This ensured that all proposed equations and answers in letters, hypotheses, and feedback on hypotheses were available to each person throughout the problem, reducing demands on memory. All group members and individuals had additional scratch paper for computations and notes. After the two problems were completed, the experimenter explained the purpose of the research to the participants, answered any questions, gave them a written debriefing with a reference for further reading, asked them not to discuss the experiment with potential future participants, and thanked them for their participation.

Results

We first determined the best, second best, third best, fourth best, and fifth best of the 5 individuals in each of the 40 replications by the number of trials over the two problems (if 2 individuals had the same number of trials, we randomly assigned them to the appropriate conditions). Nonsolvers in the allotted 10 trials were considered to require 11 trials. Means were 12.98 for the best individuals, 15.00 for the second best, 16.90 for the third best, 18.58 for the fourth best, and 19.78 for the fifth best. Because the 40 codings of letters to numbers differed in difficulty, we conducted a randomized blocks analysis of variance (ANOVA) on the number of trials to solution, with the 40 replications as a blocking variable. This analysis indicated a significant main effect of the best, second-best, third-best, fourth-best, and fifth-best individuals, $F(4, 156) = 156.68$, $p < .001$, $MSE = 0.9498$. Tukey's comparisons indicated that all 10 pairwise differences were significant ($p < .01$).

The 40 five-person groups were compared with the 40 best, second-best, third-best, fourth-best, and fifth-best individuals. Four of the 5 individuals in each replication were randomly selected, and the 40 four-person groups were compared with the 40 best, second best, third best, and fourth best of these 4 individuals. Similarly, 3 of the 5 individuals in each replication were randomly selected, and the 40 three-person groups were compared with the 40 best, second best, and third best of these 3 individuals. In addition, 2 of the 5 individuals in each replication were randomly selected, and the 40 two-person groups were compared with the 40 best and second best of these 2 individuals.

Table 4 gives the means and standard deviations for Problems 1 and 2 for number of trials to solution and letters per equation for the best, second best, third best, fourth best, and fifth best of five

Table 4
Trials to Solution and Letters per Equation for Best, Second-Best, Third-Best, Fourth-Best, and Fifth-Best Individuals

			Problem 1		Problem 2	
Group Size	Individual	Variable	*M*	*SD*	*M*	*SD*
Five	Best	Trials	6.83	1.34	6.15	1.23
		Letters per equation	2.33	0.67	2.37	0.93
	Second	Trials	7.95	1.50	7.05	1.28
		Letters per equation	2.15	0.29	2.29	0.73
	Third	Trials	8.73	1.43	8.18	1.47
		Letters per equation	2.09	0.23	2.16	0.66
	Fourth	Trials	9.73	1.11	8.85	1.44
		Letters per equation	2.11	0.35	2.15	0.40
	Fifth	Trials	10.20	1.14	9.58	1.39
		Letters per equation	2.01	0.19	2.11	0.41
Four	Best	Trials	7.15	1.48	6.40	1.36
		Letters per equation	2.20	0.50	2.34	1.04
	Second	Trials	8.23	1.56	7.53	1.41
		Letters per equation	2.15	0.32	2.31	0.76
	Third	Trials	9.33	1.21	8.70	1.49
		Letters per equation	2.04	0.08	2.09	0.29
	Fourth	Trials	10.18	1.20	9.35	1.35
		Letters per equation	2.02	0.20	2.09	0.35
Three	Best	Trials	7.30	1.44	6.40	1.26
		Letters per equation	2.32	0.67	2.36	0.94
	Second	Trials	8.68	1.42	7.83	1.69
		Letters per equation	2.17	0.43	2.30	0.80
	Third	Trials	10.08	1.16	9.15	1.51
		Letters per equation	2.01	0.20	2.11	0.41
Two	Best	Trials	8.00	1.89	7.33	1.64
		Letters per equation	2.22	0.59	2.27	0.87
	Second	Trials	9.60	1.45	8.83	1.60
		Letters per equation	2.06	0.31	2.20	0.74

individuals; best, second best, third best, and fourth best of four individuals; best, second best, and third best of three individuals; and best and second best of two individuals. Table 5 gives the means and standard deviations for five-person, four-person, three-person, and two-person groups.

Groups Versus Individuals

Table 6 gives the results of one-tailed *t* tests and standard effect sizes for comparisons of the groups and the best, second best, and so forth of an equivalent number of individuals at each group size for trials to solution and letters per equation. As indicated in Table 6, the five-person groups had significantly fewer trials to solution and more letters per equation than each of the best, second-best, third-best, fourth-best, and fifth-best individuals. Similarly, the four-person groups had significantly fewer trials to solution and more letters per equation than each of the four types of individuals, and three-person groups had significantly fewer trials to solution and more letters per equation than each of the three types of individuals. The two-person groups did not differ significantly from the best individuals and had significantly fewer trials to solution than the second-best individuals.

Group Size

A 4 (group size: two, three, four, five) × 2 (problems: one, two) randomized blocks ANOVA with repeated measures on the second variable for trials to solution indicated a significant main effect of group size, $F(3, 117) = 11.04$, $p < .001$, $MSE = 2.9565$. Tukey's pairwise comparisons indicated fewer trials to solution for each of three-person, four-person, and five-person groups than two-person groups, with nonsignificant differences among three-person, four-person, and five-person groups. The main effect of successive problems was significant, $F(1, 117) = 63.38$, $p < .001$, $MSE = 0.8721$, with fewer trials to solution on Problem 2 ($M = 5.96$) than Problem 1 ($M = 6.79$). The Size × Problems interaction was nonsignificant, $F(1, 117) = 2.04$.

A similar ANOVA for letters per equation indicated a nonsignificant main effect of group size, $F(3, 117) = 1.53$. The main effect of successive problems was significant, $F(1, 117) = 17.22$, $p < .001$, $MSE = 0.4142$, with more letters per equation on Problem 2 ($M = 2.89$) than Problem 1 ($M = 2.60$). The Size × Problems interaction was nonsignificant, $F(1, 117) < 1$.

Table 5
Trials to Solution and Letters per Equation for Five-Person, Four-Person, Three-Person, and Two-Person Groups

Group size	Variable	Problem 1 *M*	Problem 1 *SD*	Problem 2 *M*	Problem 2 *SD*
Five	Trials	6.25	1.13	5.70	1.18
	Letters per equation	2.77	1.37	3.14	1.96
Four	Trials	6.50	1.38	5.78	1.31
	Letters per equation	2.63	1.38	2.85	1.50
Three	Trials	6.45	1.52	5.65	1.41
	Letters per equation	2.71	1.38	3.16	1.84
Two	Trials	7.95	1.62	6.70	1.77
	Letters per equation	2.27	0.84	2.42	1.10

Table 6
Comparison of Groups and Equivalent Number of Individuals at Each Group Size for Trials to Solution and Letters per Equation

Group and individual	Trials *t*	Trials *SES*	Letters *t*	Letters *SES*
Five				
Best	2.72	0.69	3.35	0.47
Second	8.10	1.30	4.09	0.59
Third	13.15	1.67	4.63	0.66
Fourth	17.60	1.81	4.60	0.66
Fifth	20.79	1.87	4.99	0.73
Four				
Best	3.11	0.50	2.76	0.41
Second	8.47	1.15	3.02	0.47
Third	14.02	1.55	4.10	0.64
Fourth	17.67	1.69	4.05	0.65
Three				
Best	3.46	0.61	2.88	0.48
Second	9.53	1.29	3.40	0.59
Third	15.43	1.65	4.22	0.75
Two				
Best	1.13[a]		0.61[a]	
Second	6.24	1.10	1.31[a]	

Note. All *p*s < .001, except where marked with a superscript *a*.
[a] nonsignificant.

Regression of Trials to Solution on Letters per Equation

Regression of the total trials to solution over the two problems on the total letters per equation over the two problems indicated an *R* of .38 ($R^2 = .14$, $p < .02$) for two-person groups, *R* of .65 ($R^2 = .42$, $p < .001$) for three-person groups,; *R* of .34 ($R^2 = .11$, $p < .03$) for four-person groups, and *R* of .60 ($R^2 = .36$, $p < .001$) for five-person groups. Separate regressions for Problems 1 and 2 ranged from .18 for four-person groups on Problem 1 to .73 for three-person groups on Problem 2. The *R*s were higher on Problem 2 than Problem 1 for all group sizes, indicating a fuller realization of the effectiveness of multiletter equations on the second problem. Thus, as predicted, the number of trials to solution was influenced by the complexity of the proposed equations.

Enjoyment

After the second problem, the group members and individuals rated their enjoyment of the experiment on a 1–10 scale (10 being the highest). Ratings were averaged over the members of a given group. Both the individuals ($M = 7.96$) and the group members ($M = 7.24$) rated their enjoyment quite highly, although individuals rated their enjoyment more highly than the group members, $F(1, 334) = 19.06$, $p < .001$. Enjoyment ratings did not differ for the members of different-sized groups, $F(3, 145) < 1$. The main effect of individuals (best, etc.) was significant, $F(4, 182) = 4.42$, $p < .01$. Each of the best ($M = 8.33$), second-best ($M = 8.39$), and third-best ($M = 8.18$) individuals enjoyed the experiment more than the fifth-best individuals ($M = 7.15$). All other pairwise comparisons were nonsignificant.

Discussion

Group Versus Individual Performance

Each of the three-person, four-person, and five-person groups had significantly fewer trials to solution and more letters per equation than the best of an equivalent number of individuals. In contrast, the two-person groups did not differ significantly from the best individuals, although they had significantly fewer trials to solution than the second-best individuals. These results replicate the previous superiority of three-person groups over the best of three independent individuals (Laughlin et al., 2003) and four-person groups over the best of four independent individuals (Laughlin et al., 2002) and extend them to the superiority of five-person groups over the best of five independent individuals. To our knowledge, these three experiments are the only reports that groups of size three, four, and five perform better than the best of an equivalent number of individuals. As the best group member is comparable to the best independent individual, these results also indicate that the groups performed better than their best member would have performed alone.

We attribute this superiority of three-person, four-person, and five-person groups over the best of an equivalent number of individuals to the highly intellective nature of letters-to-numbers problems, which allow recognition and adoption of correct responses, recognition and rejection of erroneous responses, and effective collective information processing (Hinsz, Tindale, & Vollrath, 1997; Laughlin, VanderStoep, & Hollingshead, 1991). Letters-to-numbers problems strongly fulfill Laughlin and Ellis's (1986) four conditions of demonstrability. First, the group members understand and agree on the underlying conceptual systems of arithmetic, algebra, and logic. Second, there is sufficient information to demonstrate the superiority of strategies such as multiletter equations relative to obvious but less effective two-letter substitution strategies and to demonstrate and reject erroneous inferences. Virtually any sequence of equations of any degree of complexity contains some information, and the groups were able to process this information more effectively than the best individuals. Third, the members who had not considered effective strategies and reasoning recognized the effectiveness when such strategies were proposed by other members. Fourth, the members who proposed the effective strategies and reasoning had the ability, motivation, and time to demonstrate the effectiveness to the other members. Thus, the group members combined their abilities and resources to perform better than the best of an equivalent number of individuals on the highly intellective complementary group task (Steiner, 1966).

Tindale and Kameda (2000) and Kameda, Tindale, and Davis (2003) generalized the first of these conditions in their concept of *social sharedness*, the degree to which preferences and cognitions are shared among group members at the outset of group interaction. Building on these shared preferences (the objective of solving in as few trials as possible and the norms of interpersonal influence, e.g., accepting a demonstrably effective strategy) and cognitions (the conceptual systems and operations of arithmetic, algebra, and logic), the groups combined the abilities, skills, and insights of their members and thus performed better than the best of an equivalent number of individuals.

Group Size

Each of the three-person, four-person, and five-person groups had fewer trials to solution and proposed more complex equations than the two-person groups. The three-person, four-person, and five-person groups did not differ significantly from each other on either trials to solution or letters per equation. As there were 40 replications of each group size, we may be confident that these nonsignificant differences between group sizes three, four, and five were not due to insufficient statistical power.

Our review of previous research suggests that performance improves with increasing group size for problems of moderate difficulty that require understanding of verbal, quantitative, or logical conceptual systems, but performance has not been shown to improve with increasing group size for eureka problems. The current finding that three-person groups performed better than two-person groups is consistent with this research, as letters-to-numbers problems require understanding of arithmetic, algebra, and logic and systematic reasoning over a series of trials rather than a single insight, and they are of moderate but not excessive difficulty for intelligent and motivated college students.

Why was there no further improvement as group size increased from three to four to five? We have suggested that the crucial aspect of effective performance on letters-to-numbers problems is the use of multiletter equations to identify two or more letters per equation rather than the use of a simple but inefficient two-letter substitution strategy that identifies only one letter per trial. Assuming the individual probability of using two-letter equations on all trials in previous research and assuming that multiletter equations are demonstrably preferable to two-letter equations if proposed by at least one group member predicts progressively smaller improvement with group size beyond three members. Coordination difficulties (Steiner, 1972) and production blocking (Diehl & Stroebe, 1987, 1991; Valacich, Dennis, & Connolly, 1994) might also have increased with increasing group size beyond three members. Although these processes should be minimal in three-person groups, they may be more detrimental in four-person and five-person groups. In contrast, we do not believe that there was an appreciable motivation loss as a result of free riding (Kerr, 1983; Olson, 1965) from three-person groups to four-person and five-person groups, as the group members rated the experiment as quite enjoyable and there were nonsignificant differences for group size.

Thus, three group members were necessary and sufficient for the groups to perform better than the best of an equivalent number of independent individuals. If groups of size three perform as well as groups of larger size, it is obviously a more efficient use of human and logistic resources to use three-person groups. Further research should be conducted to determine whether three persons are necessary and sufficient for groups to perform better than the best of an equivalent number of individuals on other problem-solving tasks, such as survival problems (e.g., Littlepage, Schmidt, Whisler, & Frost, 1995). The research program on team performance on realistic command and control tasks of Ilgen, Hollenbeck, and colleagues has used four-person groups (e.g., Hedlund, Ilgen, & Hollenbeck, 1998; see Ilgen, Hollenbeck, Johnson, & Jundt, 2005, for a review), and it is theoretically and practically important to determine whether group performance is comparable for smaller groups of size three and improves with groups of size five and larger.

Future Research: Training, Transfer, and Expertise

Letters-to-numbers problems combine aspects of hypothesis testing (e.g., Klayman & Ha, 1987), mathematical and logical reasoning (e.g., Laughlin & Ellis, 1986; Stasson, Kameda, Parks, Zimmerman, & Davis, 1991), cryptographic reasoning (e.g., Newell & Simon, 1972; Singh, 1999), and collective induction (e.g., Crott, Giesel, & Hoffman, 1998; Laughlin, 1999). These correspondences indicate that letters-to-numbers problems are a useful and interesting domain for future research on the interrelated issues of training, transfer, and expertise in small-group problem solving (e.g., Bonner, 2004; Bonner, Baumann, & Dalal, 2003; Hollingshead, 1998; Liang, Moreland, & Argote, 1995; Littlepage et al., 1995). For example, individuals could be trained to use effective multiletter substitution or known answer strategies (see Laughlin et al., 2003, for further discussion of effective strategies) before assembling for further group problem solving. Through such member training, groups may be able to overcome the common knowledge (e.g., Gigone & Hastie, 1993) and hidden profile (e.g., Stasser & Stewart, 1992) effects, whereby groups discuss common initial preferences and shared information, fail to discuss unique unshared information, and hence make suboptimal decisions. In contrast to this research, in which a single member with critical information may not be able to convince other members of the validity of the information, letters-to-numbers problems enable a member who has been trained on an effective strategy to demonstrate the effectiveness of the strategy to the other group members. Similarly, letters-to-numbers problems are a useful and interesting task for future research on group-to-individual transfer: Does effective group problem solving transfer to subsequent individual problem solving?

Conclusions

Groups of size three, four, and five performed better than the best of an equivalent number of individuals, but groups of size two performed at the level of the best of two individuals. Groups of size three, four, and five performed better than groups of size two but did not differ from each other. These results suggest that groups of size three are necessary and sufficient to perform better than the best of an equivalent number of individuals on intellective problems.

References

Bonner, B. L. (2004). Expertise in group problem solving: Recognition, social combination, and performance. *Group Dynamics: Theory, Research, and Practice, 4,* 277–290.

Bonner, B. L., Baumann, M. R., & Dalal, R. (2003). The effects of variability and expectation on utilization of member expertise and group performance. *Organizational Behavior and Human Decision Processes, 88,* 719–736.

Bray, R. M., Kerr, N. L., & Atkin, R. S. (1978). Effects of group size, problem difficulty, and sex on group performance and member reactions. *Journal of Personality and Social Psychology, 36,* 1224–1240.

Crott, H. W., Giesel, M., & Hoffman, C. (1998). The process of inductive inference in groups: The use of positive and negative hypothesis and target testing in sequential rule-discovery tasks. *Journal of Personality and Social Psychology, 79,* 938–952.

Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology, 53,* 497–509.

Diehl, M., & Stroebe, W. (1991). Productivity loss in idea-generating groups: Tracking down the blocking effect. *Journal of Personality and Social Psychology, 61,* 392–403.

Gigone, D., & Hastie, R. (1993). The common knowledge effect: Information sharing and group judgment. *Journal of Personality and Social Psychology, 65,* 959–974.

Hastie, R. (1986). Review essay: Experimental evidence on group accuracy. In G. Owen & B. Grofman (Eds.), *Information pooling and group accuracy* (pp. 129–157). Westport, CT: JAI Press.

Hedlund, J., Ilgen, D. R., & Hollenbeck, J. R. (1998). Decision accuracy in computer-mediated versus face-to-face decision-making teams. *Organizational Behavior and Human Decision Processes, 76,* 30–47.

Hill, G. W. (1982). Group versus individual performance: Are $N + 1$ heads better than one? *Psychological Bulletin, 91,* 517–539.

Hinsz, V. B., Tindale, R. S., &Vollrath, D. A. (1997). The emerging conceptualization of groups as information processors. *Psychological Bulletin, 121,* 43–64.

Hollingshead, A. B. (1998). Group and individual training: The impact of practice on performance. *Small Group Research, 29,* 254–280.

Ilgen, D. R., Hollenbeck, J. R., Johnson, M., & Jundt, D. (2005). Teams in organizations: From input-output models to IMOI models. *Annual Review of Psychology, 56,* 517–543.

Kameda, T., Tindale, R. S., & Davis, J. H. (2003). Cognitions, preferences, and social sharedness: Past, present, and future directions in group decision making. In S. L. Schneider & J. Shanteau (Eds.), *Emerging perspectives on judgment and decision research* (pp. 458–485). Cambridge, England: Cambridge University Press.

Kerr, N. L. (1983). Motivation losses in task-performing groups: A social dilemma analysis. *Journal of Personality and Social Psychology, 45,* 819–828.

Kerr, N. L., & Tindale, R. S. (2004). Group performance and decision making. *Annual Review of Psychology, 56,* 623–655.

Klayman, J., & Ha, Y.-M. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review, 94,* 211–228.

Laughlin, P. R. (1999). Collective induction: Twelve postulates. *Organizational Behavior and Human Decision Processes, 80,* 50–69.

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes, 88,* 605–620.

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology, 22,* 177–189.

Laughlin, P. R., Kerr, N. L., Davis, J. H., Halff, H. M., & Marciniak, K. A. (1975). Group size, member ability, and social decision schemes on an intellective task. *Journal of Personality and Social Psychology, 31,* 522–535.

Laughlin, P. R., VanderStoep, S. W., & Hollingshead, A. B. (1991). Collective versus individual induction: Recognition of truth, rejection of error, and collective information processing. *Journal of Personality and Social Psychology, 61,* 50–67.

Laughlin, P. R., Zander, M., L., Knievel, E. M., & Tan, T. K. (2003). Groups perform better than the best individuals on letters-to-numbers problems: Informative equations and effective reasoning. *Journal of Personality and Social Psychology, 85,* 684–694.

Levine, J., & Moreland, R. L. (1998). Small groups. In D. T. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *The handbook of social psychology* (Vol. 2, pp. 415–467). New York: McGraw-Hill.

Liang, D. W., Moreland, R., & Argote, L. (1995). Group versus individual training and group performance: The mediating role of transactive memory. *Personality and Social Psychology Bulletin, 21,* 384–393.

Littlepage, G. E., Schmidt, G. W., Whisler, E. W., & Frost, A. G. (1995). An input-process-output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology, 69,* 877–889.

Lorge, I., & Solomon, H. (1959). Individual performance and group performance in problem solving related to group size and previous exposure to the problem. *Journal of Psychology, 48,* 105–114.

Lorge, I., & Solomon, H. (1960). Group and individual performance in problem solving related to previous exposure to problem, level of aspiration, and group size. *Behavioral Science, 5,* 28–39.

Luchins, A. S. (1942). Mechanization in problem solving: The effects of Einstellung. *Psychological Monographs,* 54(6, Whole No. 248).

Maier, N. R. F., & Solem, A. R. (1952). The contribution of a discussion leader to the quality of group thinking: The effective use of minority opinions. *Human Relations, 5,* 277–288.

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Englewood Cliffs, NJ: Prentice Hall.

Olson, M. (1965). *The logic of collective action.* Cambridge, MA: Harvard University Press.

Singh, S. (1999). *The code book.* New York: Doubleday.

Stasser, G., & Stewart, D. D. (1992). Discovery of hidden profiles by decision-making groups: Solving a problem versus making a judgment. *Journal of Personality and Social Psychology, 63,* 426–434.

Stasson, M. F., Kameda, T., Parks, C. D., Zimmerman, S. K., & Davis, J. H. (1991). Effects of assigned group consensus requirements on group problem solving and group members' learning. *Social Psychology Quarterly, 54,* 25–35.

Steiner, I. D. (1966). Models for inferring relationships between group size and potential group productivity. *Behavioral Science, 11,* 273–283.

Steiner, I. D. (1972). *Group process and productivity.* New York: Academic Press.

Taylor, D. W., & Faust, W. L. (1952). Twenty questions: Efficiency in problem-solving as a function of size of group. *Journal of Experimental Psychology, 44,* 360–368.

Terman, L. M. (1956). *Manual for Concept Mastery Test.* New York: Psychological Corporation.

Thomas, E. J., & Fink, C. F. (1961). Models of group problem solving. *Journal of Abnormal & Social Psychology, 63,* 53–63.

Thorndike, R. L. (1937). The effect of discussion upon the correctness of group decisions, when the factor of majority influence is allowed for. *Journal of Social Psychology, 9,* 343–362.

Tindale, R. S., & Kameda, T. (2000). "Social sharedness" as a unifying theme for information processing in groups. *Group Processes and Intergroup Relations, 3,* 123–140.

Valacich, J. S., Dennis, A. R., & Connolly, T. (1994). Idea-generation in computer-mediated groups: A new ending to an old story. *Organizational Behavior and Human Performance, 57,* 448–467.

Appendix

Instructions for Letters-to-Numbers Problems

This is an experiment in problem solving. The objective is to figure out a code in as few trials as possible. The numbers 0–9 have been coded as the letters *A*–*J* in some random order. You will be trying to find out which letter corresponds to which number. It is important to remember that all we are doing is changing the characters used to represent the numbers. We are not changing the way that the number system works. That is, we are still using the same decimal number system you have been using all of your life. Below is an example of a random code:

$A = 3, B = 5, C = 8, D = 2, E = 1, F = 6, G = 4, H = 7, I = 0, J = 9$

First you will come up with addition or subtraction equations using the letters *A*–*J* that will be solved by the experimenter who will give you the answer in letter form. Then you will make a guess as to what one of the letters represents, and the experimenter will tell you whether or not the guess is correct (True) or incorrect (False). When you feel you know the full coding of letters to numbers, propose it in the space provided on your data sheet. When you have correctly mapped all ten letters to all ten numbers, you will have solved the problem. Remember, the objective is to solve the problem in as few trials as possible.

Here are four example trials using the random code above. Note that underlined letters represent experimenter feedback.

Trial	Equation	Hypothesis	Feedback
1	$A + B = \underline{C}$	$A = 1$	False
2	$B + C = \underline{EA}$	$A = 8$	False
3	$F + A + D = \underline{EE}$	$E = 1$	True
4	$H - J = \underline{-D}$	$I = 0$	True

On the first trial the person chooses the equation $A + B = ?$, and the experimenter tells the person that the solution to this equation is *C*. This is because $A = 3$ and $B = 5$, which sums to 8, the letter represented by *C*. The person then guesses that *A* represents the number 1, and the experimenter indicates that this is not the case, or False. On the second trial the person asks the solution to the equation $B + C = ?$ and is told that the answer is *EA*. This is because $B = 5$ and $C = 8$, which sums to 13 or *EA*. Note that on the third trial the person chooses to add three letters together. You may use as many letters as you desire in your equations. Note that on the fourth trial the person chooses to use a subtraction equation. You may use either addition or subtraction equations as you see fit.

Received June 8, 2004
Revision received July 22, 2005
Accepted August 3, 2005 ■