



Supporting Online Material for
**Evidence for a Collective Intelligence Factor
in the Performance of Human Groups**

Anita Williams Woolley,* Christopher F. Chabris, Alexander Pentland,
Nada Hashmi, Thomas W. Malone

*To whom correspondence should be addressed. E-mail: awoolley@cmu.edu

Published 30 September 2010 on *Science* Express
DOI: 10.1126/science.1193147

This PDF file includes:

Materials and Methods
Tables S1 to S4
References

Correction: The labeling in Table S4 was corrected.

Supporting Online Material

Materials and Methods

Study 1

Participants

Forty teams of three participants were recruited from the general public in the Boston area via internet advertising and invited to the laboratory for a session of up to five hours. Fifty-one percent of the participants were male, and they ranged in age from 18 to 66 ($M=32.1$, $SD=12.33$). All participants were paid for their participation.

Procedure

When participants arrived at the laboratory they were randomly assigned to participate in a team or, in the event of overscheduling, randomly selected for exclusion. Before meeting their teams, individuals completed a brief intelligence test (described below). Following the intelligence test, teams began working together on the group tasks in a private laboratory room.

Tasks and Measures

Individual intelligence test. Raven's Advanced Progressive Matrices (RAPM) is a standardized test of general fluid reasoning capacity. Participants were given 10 minutes to complete half of the 36-item Raven's II test (all of the odd-numbered items). Each item presents a 3 x 3 array (or "matrix") of shapes with the lower-right corner empty. Based on patterns in the sequences of shapes, subjects must pick which of 8 other shapes properly belongs in the empty space. There is only one correct answer per question. Questions become progressively more difficult as the test goes on. Examples of items are available from Harcourt Education.

Group tasks. We selected tasks from the McGrath Task Circumplex (*I*), an established and validated taxonomy characterizing tasks according to the dominant coordination process required for its accomplishment by a group. The Taxonomy identifies four main types of tasks: (1) Quadrant I includes "Generate" tasks which include brainstorming tasks and anything involving the development of new ideas or information; (2) Quadrant II includes "Choose" tasks which involve deciding about issues that either have a correct answer or which are matters of judgment, with some research noting important distinctions among intellectual and judgmental tasks (2); (3) Quadrant III includes "Negotiate" tasks which involve resolving conflicts of interest or points of view; and (4) Quadrant IV includes "Execute" tasks which involve performances and psycho-motor tasks. We included at least one task from each quadrant. The tasks are described below. Descriptive statistics and intercorrelations of measures are shown in Table 1.

Brainstorming(Quadrant I). Groups spent 10 minutes brainstorming possible uses for a brick. Groups received one point for each non-redundant idea they generated, independent of quality of the ideas.

Group Matrix Reasoning (Quadrant II). Groups completed the even-numbered questions of RAPM questions as a group. Groups were scored on the number of items answered correctly.

Group Moral Reasoning (Quadrant II). Using the “Disciplinary Action Case” (3), groups decided on disciplinary actions in a fictitious case in which a college basketball player bribed an instructor to change his grade on an exam. The groups were given a list of five issues having to do with how to treat the athlete and the instructor. Issues included what to do about the student’s grade in the course and whether to suspend him from school. The groups' task was to select one choice from a list of alternatives provided for each issue. In addition, groups were told to take into account the conflicting interests of the faculty, college administration, and the athletic department when making their decisions. Responses were scored using a rubric that reflected the degree to which the groups considered the balance of competing perspectives on the problem.

Plan shopping trip (Quadrant III). The groups' task was to plan a shopping trip as if they were all residents of the same house sharing a single car. Each group member was given a different list of groceries they needed for the week, and various constraints applied. For instance, there were better and worse places to buy the different items, with cheaper and higher quality options requiring more driving time. Maps were provided with information on distances and time for reaching each store. Solutions were scored as follows: (a) each item purchased = +1 point, (b) bonus for high quality item = +1 point, (c) bonus for lower priced alternative = +2 points, (d) penalty for leaving frozen items in the car beyond 30 minutes = loss of all points for that item. The groups' goal was to work out a plan in which they could purchase as many high-value items from each of their lists as possible and thus earn as many points as possible for their group.

Group typing (Quadrant IV). Groups were provided a hard copy of a complicated text and worked for 10 minutes to simultaneously type as much of the text as possible into a shared online document. Participants were each seated in front of a separate computer and worked in the shared online document where they could see each others' work with a slight delay. Teams earned one point for each word correctly typed, and lost one point for each word skipped and for each typo. Team members thus needed to carefully coordinate their work to avoid typing over the work of other members or skipping whole sections (which would result in the loss of many points).

Criterion task: Video checkers. Group members were seated together in front of a single computer to play checkers against a computerized opponent. Members were first familiarized with the game and the rules individually and given 5 minutes to talk about their strategy. Teams then played a practice match and then one test match against the computer opponent. Only the results of the test match were used for the analysis. Teams received one point for each move they made, two points for each piece captured, and 3 points for each king they earned.

Correlations, Factor Analysis, and OLS Regression Results

The full table of correlations among all variables measured is available in Table 1 of the main article. The initial eigenvalues associated with each factor and after principal axis factor analysis are displayed in Table S1a, and the loadings associated with each of the group tasks are displayed in Table S1b. Results from OLS Regression, using average individual intelligence and the *c*-factor as predictors of teams' video checkers scores are in Table S2.

Study 2

Participants

A total of 579 individuals working in 152 teams were recruited from the general public in the Boston and Pittsburgh areas via internet advertising, and invited to the laboratory for a session of up to four hours. The samples from Boston and Pittsburgh did not vary significantly in terms of gender composition (40% male in Boston vs. 38% male in Pittsburgh), age ($M=29.74$, $SD=6.01$ in Boston vs. $M=23.73$, $SD=3.12$ in Pittsburgh) or average individual intelligence as measured by the Wonderlic Personnel Test ($M=22.9$, $SD=6.81$ in Boston vs. $M=24.4$, $SD=7.07$ in Pittsburgh). As in Study 1, all participants were paid for their participation.

Procedure

When participants arrived at the laboratory they were randomly assigned to work by themselves or in a team of 2, 3, 4, or 5 members. Before meeting their teams, individuals completed measures of their individual intelligence, personality, and social sensitivity (described below). Following these measures, teams began working together on the group tasks in a private laboratory room.

Tasks and Measures

In this study, all groups performed the same five group tasks as in Study 1, and the groups at the Pittsburgh site also performed five additional tasks (see below). Also, as described below, we tested a different criterion task, used an alternative measure of individual intelligence, and collected additional measures of individual differences and group climate. Descriptive statistics and correlations for all these measures are shown in Tables S3a and S3b.

Individual Measures. Participants completed individual intelligence, personality, and social sensitivity tests prior to working with their group, and measures of group satisfaction, motivation, social cohesiveness, and psychological safety after the group work period was over.

Individual intelligence. Individuals completed the Wonderlic Personnel Test (WPT) (4), a 50-item measure that is widely used for assessing individual intelligence. The WPT includes verbal, mathematical, and spatial items, and studies comparing it to longer and more variegated measures of intelligence, such as the WAIS, find an average correlation of .92 between the scores on the two tests (4-7). Individuals earned one point for each item answered correctly, and scores were averaged for the group as a whole.

Individual Personality. Participants completed the NEO Five Factor Inventory (8), one of the most widely used personality tests. This test measures the five primary domains of adult personality: Extraversion, Agreeableness, Conscientiousness, Openness to Experience, and Neuroticism. Sixty items are responded to on a 1 to 5 scale; the mean for each scale was calculated for each participant. Scores on each dimension were averaged for the group as a whole.

Social Sensitivity. Participants completed the "Reading the Mind in the Eyes" test (9). In this test, a participant is presented with a series of 35 photographs of the eye-region of the face of different actors and actresses, and is asked to choose which of four words best describes what the person in the photograph is thinking or feeling. It is considered an 'advanced theory of mind test' which gauges the ability to attribute mental states to oneself or another person. In the area of psychometrics, our sample was similar in mean and SD on the Eyes test to the original general population sample of Baron-Cohen et al. (9)(ours: $M=25.9$, $SD=2.8$; original: $M=26.2$, $SD=3.6$) and we found, like Baron-Cohen et al., that women scored slightly higher than men. The test has been shown to have adequate test-retest reliability (10) and to consistently differentiate control subjects from autism-spectrum individuals, who are below-average in social sensitivity (9). The Eyes test has been shown to be sensitive in children to levels of prenatal testosterone (11) and in adults to administration of oxytocin (12). This suggests that the Eyes test is measuring a fundamental property of individual brain function; if performance on the test were completely determined by contextual factors then we would not expect to see such correlations or experimental effects. Individual scores were averaged for the group as a whole.

Group satisfaction was measured using a three-item scale developed and validated in prior research (13). Participants indicated their agreement to statements such as "I have been very satisfied working on this team" on a five-point scale. Team members' responses demonstrated adequate reliability for analysis at the team level ($ICC_1 = .50$, $p<.05$; $ICC_2=.75$, $r_{wg}=.92$).

Motivation was measured using a four-item scale developed and validated in prior research (13). Participants indicated their agreement to statements such as "I would feel bad and unhappy if our team has performed poorly" (reverse-scored) on a five-point scale. Team members' responses demonstrated adequate reliability for analysis at the team level ($ICC_1 = .53$, $p<.05$; $ICC_2=.72$, $r_{wg}=.90$).

Social cohesiveness was measured using a three-item scale adapted from prior research (14). Participants indicated their agreement to statements such as "Members of this group would enjoy being at a party together" on a five-point scale. Team members' responses demonstrated adequate reliability for analysis at the team level ($ICC_1 = .59$, $p<.01$; $ICC_2=.77$, $r_{wg}=.91$).

Psychological Safety was measured using an eight-item scale developed and validated by prior research(15). Participants indicated their agreement to statements such as "It is difficult to ask other members of this team for help" on a five-point scale. Team members' responses demonstrated adequate reliability for analysis at the team level ($ICC_1 = .65$, $p<.01$; $ICC_2=.82$, $r_{wg}=.93$).

Group tasks and measures. Groups worked together on all of the tasks described for Study 1, and a subset of 107 groups worked on five additional tasks (described below). As they worked, a subset of 46 groups wore sociometric badges capturing their communication behavior, as described in more detail below.

Word completions (Quadrant I). Groups were given 10 minutes to come up with as many words beginning with "s" and ending with "n" as possible. Groups earned 1 point for each non-redundant English word.

Spatial problems (Quadrant II). Groups were given 10 minutes to figure out as many ways as they could for fitting 6 three-dimensional rectangles into a three dimensional container. Groups indicated their solutions by filling in dotted lines on a diagram provided. Groups were given 1 point for each correct solution that used the same rule (e.g. placing them all in the same direction) and 2 points for each correct solution that used a new rule (e.g. changing the orientations of some of the pieces).

Incomplete words (Quadrant II). Groups were given a set of 36 words with 2-3 letters missing (ex. " _ ech_ _ que" would be "technique"). Groups had 10 minutes to complete as many words as possible. Groups earned 1 point for each correctly spelled English word they completed.

Estimation Problems (Quadrant II). Groups were given a set of 20 items requiring estimation of a quantity (e.g., "What was the median age in the U.S. in 2009?") and had 10 minutes to answer as many as possible. Groups were scored based on the absolute percent deviation of their estimate from the correct answer.

Reproducing art (Quadrant IV). Groups were given a hard copy of a picture created by coloring cells in a spreadsheet, and had to duplicate the picture as accurately as possible using a shared online spreadsheet tool. Groups had 5 minutes to explore the tool, followed by one simple five-minute training task, and then were given 10 minutes to complete the trial task. Groups were given 1 point for each cell colored correctly.

Criterion task: Architectural design. Groups were given a complex multi-faceted task in which they needed to design and build a house, garage, and pool out of a limited set of building blocks, developed by Woolley (16). Groups' structures were also expected to conform to a number of strict building codes and other criteria. Final structures were scored on the basis of their size, aesthetic quality, and durability but could receive large penalties for violating requirements such as the height of walls, and the size of door and window openings. Teams received 15 minutes of instructions, 10 minutes of planning time, and 20 minutes to build their structures.

Speaking turn variance was calculated using Sociometric Badge technology (17). Participants in 46 groups wore small boxes on a cord around their neck which recorded audio data from their communication. An in-house algorithm processed the raw audio data to differentiate each individual's voice from other voices on the basis of the audio volume, pitch, and prosody of their voice, and the individual's silence between sentences when others

participants were speaking was equated to the beginning of a new speaking turn. The program then calculated how many times an individual spoke over the course of the group's work together. The complete interaction and variance pattern was derived by aggregating these indices across all individuals in the group to come up with a single index of speaking turn variance for the group.

Correlations, Factor Analysis, and Regression Results

The full table of correlations among all variables measured is shown in Table S3b. Both initial eigenvalues and those associated with each factor extracted via principal axis factor analysis are displayed in Table S1, and the loadings associated with each of the group tasks are displayed in Table S1b. Results from OLS regression, using average individual intelligence and groups' *c*-factor as predictors of teams' performance on the architectural design task are displayed in Table S2. Table S4 displays results from OLS regression analysis examining the combined influence of the percent of females, social sensitivity, and conversational turn taking on *c*.

Table S1a

Eigenvalues and Percent Variance Associated with Each Component in Initial Solution and After Principal Axis Extraction

			Factor				
			1	2	3	4	5
Study 1	Initial	Total	2.17	.91	.85	.62	.45
	Eigenvalues	% Variance	43.39	18.18	16.93	12.46	9.04
	Principal Axis	Total	1.56				
	Extraction	% of Variance	31.29				
Study 2	Initial	Total	2.20	1.03	.78	.62	.37
	Eigenvalues	% Variance	44.07	20.54	15.57	12.44	7.39
	Principal Axis	Total	1.76	.35			
	Extraction	% of Variance	35.12	6.72			
Study 2 Extended Task Sample	Initial	Total	2.79	1.33	1.24	1.01	.92
	Eigenvalues	% Variance	27.86	13.35	12.40	10.08	9.24
	Principal Axis	Total	2.39	.76	.61	.57	
	Extraction	% Variance	23.87	7.56	6.08	5.67	

Table S1b

Comparison of Principal Axis Factor Loadings for Tasks in Study 1 and Study 2

	Study 1	Study 2
Brainstorming	0.33	0.72
Group Matrix Reasoning	0.74	0.80
Group Moral Reasoning	0.36	0.10
Plan Shopping Trip	0.57	0.48
Group Typing	0.68	0.61
Eigenvalue	2.17	2.21
Coefficient of congruence		0.93

* Correlation is significant at $p < .05$, two-tailed ** Correlation is significant at $p < .001$, two-tailed

Table S2

Results of OLS Regression Analyses of Effects of Average Member Intelligence and Collective Intelligence on Criterion Tasks in Study 1 and Study 2

	Study 1: Video Game (n=40)			Study 2: Architectural Design (n=152)			
	Step 1	Step 2	Step 3	Step 1	Step 2	Step 3	Step 4
Number of members ^a				-0.04	-0.03	-0.20	-0.27*
Average Member Intelligence	0.18	0.08			0.18	0.05	
Maximum Member Intelligence			0.01				0.12
Collective Intelligence		0.51**	0.53**			0.36*	0.37*
F	1.21	7.14**	6.94**	0.20	2.62	6.59**	6.58**
R ²	0.03	0.28**	0.27**	0.04	0.16	0.34**	0.35**
change R ²		0.25**	0.24**		0.12	0.18*	0.19*

* significant at $p < .05$, two-tailed ** significant at $p < .001$, two-tailed

^a Number of members is constant for Study 1 and thus not a variable in the analysis

Table S3a
Descriptive Statistics and Correlations for Study 2 Tasks (n=152)

	Minimum	Max	Mean	SD
Collective intelligence (c)	-3.21	3.31	0.00	1.00
Brainstorming	0.00	29.00	13.63	4.48
Group Matrix Reasoning	1.00	12.00	7.46	2.13
Group Moral Reasoning	0.00	84.00	58.89	13.54
Group Typing	7.00	1993.00	378.45	253.97
Plan Shopping Trip	-3.40	86.60	29.00	18.70
Word Completions^b	9	37	20.44	6.16
Spatial Problems^b	1	28	15.28	6.32
Incomplete Words^b	22	50	47.91	4.71
Estimation Problems^b	14.34	79.76	50.69	13.45
Reproducing Art^b	56	1506	1192.96	250.98
Satisfaction	2.75	4.67	3.50	0.29
Psychological Safety	2.89	5.00	3.76	0.29
Cohesion	2.30	4.00	3.00	0.29
Motivation	2.80	4.50	3.68	0.33
Neuroticism	1.50	4.50	2.79	0.47
Extraversion	1.83	4.33	3.40	0.37
Openness to experience	2.50	4.58	3.57	0.39
Agreeableness	2.08	4.67	3.54	0.38
Conscientiousness	2.00	4.92	3.58	0.43
Social Sensitivity	18.50	35.00	25.91	2.83
Percent Female	0	100	50.16	37.22
Speaking Turn Variance^a	0.71	354.97	112.37	97.18
Maximum Member Intelligence	14.00	39.00	28.60	4.62
Average Member Intelligence	8.00	39.00	23.89	6.04
Architectural Design Task	-23020	39450	12805	9384

a. Only a portion of the sample was recorded using the sociometric badge technology (n=46)

b. Only a portion of the sample completed the extended set of problem (n=107)

Table S3b
Descriptive Statistics and Correlations for Study 2 Tasks (n=152)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
1. Collective Intelligence (c)																										
2. Brainstorming	.56**																									
3. Group Matrix Reasoning	.86*	.30*																								
4. Group Moral Reasoning	.11	.13	.04																							
5. Plan Shopping Trip	.32*	.17	.08	-.04																						
6. Group Typing	.70**	.18	.42**	.08	.03																					
7. Word Completions^b	.60**	.44**	.43**	.07	.24*	.44**																				
8. Spatial Problems^b	.34*	.36**	.34*	.15	.04	.10	.11																			
9. Incomplete Words^b	.28*	.31*	.25*	.00	.17	.09	.39**	.02																		
10. Estimation Problems^b	.20*	.29*	.18	.00	.10	-.02	.24*	.27*	.18																	
11. Reproducing Art^b	.25*	.15	.27*	.01	-.09	.18	.24*	.13	.28*	.02																
12. Satisfaction	-.07	.02	-.09	.12	-.12	-.01	-.18	-.08	-.11	-.10	-.06	(.75) ^c														
13. Psychological Safety	.08	.11	.12	.07	-.04	-.05	-.11	.08	.09	-.15	.00	.28*	(.82) ^c													
14. Cohesion	-.12	-.10	-.09	.07	.10	-.16	-.11	-.11	-.03	-.13	.02	.09	-.16	(.77) ^c												
15. Motivation	-.01	-.07	.03	.19	-.04	-.01	-.10	.02	-.02	-.10	-.14	.19	.45*	.08	(.72) ^c											
16. Neuroticism	-.09	-.15	-.04	-.01	-.07	-.04	-.21*	.04	-.15	-.07	.06	.14	-.16	.05	-.04	(.87) ^d										
17. Extraversion	-.01	.19	-.01	.08	.04	-.15	-.01	-.06	.02	-.04	-.22*	.04	.17	.06	.20*	-.32*	(.76) ^d									
18. Openness to experience	.08	.29*	.06	.00	.19	-.18	.01	.18	-.03	.04	.03	.02	.09	-.09	-.08	.07	.01	(.81) ^d								
19. Agreeableness	.01	.09	.01	-.03	.16	-.12	.02	-.13	.22*	.05	-.08	-.06	.38*	-.04	.26*	-.29*	.40**	.09	(.92) ^d							
20. Conscientiousness	-.14	-.03	-.02	.10	.06	-.12	-.32*	-.13	-.15	.00	-.09	.07	.09	.03	.24*	-.30*	.29*	-.18	.24*	(.84) ^d						
21. Social Sensitivity^e	.26*	.23*	.13	.05	.27*	.01	.13	.21*	.20*	-.05	.13	.00	.21*	-.03	.12	-.04	.03	.37*	.22*	.06						
22. Percent Female^e	.23*	-.03	.18	-.13	.03	.26*	-.10	-.06	.16	-.28*	.02	.02	-.06	-.03	.14	.10	.08	-.09	.18	.13	.19					
23. Speaking Turn Variance^{a,e}	-.41**	.21*	-.36*	.28*	-.20*	-.23*	^a	^a	^a	^a	^a	-.21*	.00	.14	^a	.02	.04	.17	-.31*	-.04	-.23*	-.23*				
24. Avg Member Intelligence	.28*	.24*	.38**	.16	.08	.09	.27*	.45**	.13	.35*	.31*	-.01	.13	-.07	-.06	.12	-.18	.25*	.01	-.25*	.35*	-.14	-.12			
25. Max Member Intelligence	.33*	.22*	.31*	.16	.18	.13	.40**	.32*	.23*	.22*	.37**	-.14	-.04	-.10	-.09	.02	-.18	.06	-.07	-.14	.30*	-.17	-.29	.64**		
26. Architectural Design Task	.28*	.21*	.22*	.04	.22*	.14	.12	.34*	.23*	.09	.32*	.05	.19	-.09	-.01	-.08	-.02	.13	.10	-.22*	.25*	.08	-.04	.18	.10	

* Correlation is significant at $p < .05$, two-tailed ** Correlation is significant at $p < .001$, two-tailed

a. Only a portion of the sample was recorded using the sociometric badge technology (n=46)

b. Only a portion of the sample completed the extended set of problem (n=107)

c. Intra-class Correlation Coefficients for associated measures

d. Cronbach's Alpha Reliability coefficient for associate measures

e. Correlations calculated while controlling for group size

Table S4

Results of OLS Regression Analyses of Effects of Percent Female, Average Member Social Sensitivity and Speaking Turn Variance on Collective Intelligence (n=46 groups)

	Collective Intelligence (c)			
	Step 1	Step 2	Step 3	Step 4
Number of members	0.09	0.19	0.21	0.28
Percent Female		0.40*	0.26	0.25
Social Sensitivity			0.37*	0.33*
Speaking Turn Variance				-0.27
F	0.28	3.01	3.91*	3.87*
R ²	0.08	0.16	0.27	0.34
change R ²		0.08	0.11*	0.07

* Coefficient is significant at $p < .05$, two-tailed

References

1. J. E. McGrath, *Groups: Interaction and Performance* (Prentice-Hall, Englewood Cliffs, NJ, 1984).
2. J. Larson, *In search of synergy in small group performance* (Psychology Press, 2009).
3. S. G. Straus, *Small Group Research* **30**, 166-187 (1999).
4. E. F. Wonderlic, C. I. Rowland, *The Journal of Applied Psychology* **23** (1939).
5. C. B. Dodrill, *Journal of consulting and clinical psychology* **49**, 668-673 (1981).
6. C. B. Dodrill, M. H. Warner, *Journal of consulting and clinical psychology* **56**, 145-147 (1988).
7. C. B. Dodrill, *Journal of consulting and clinical psychology* **51**, 316-317 (1983).
8. R. R. McCrae, P. T. Costa, *Journal of Personality and Social Psychology* **52**, 81-90 (1987).
9. S. Baron-Cohen, S. Wheelwright, J. Hill, Y. Raste, I. Plumb, *The Journal of Child Psychology and Psychiatry and Allied Disciplines* **42**, 241-251 (2001).
10. M. U. Hallerback, T. Lugnegard, F. Hjarthag, C. Gillberg, *Cognitive Neuropsychiatry* **14**, 127-143 (2009).
11. E. Chapman et al., *Social Neuroscience* **1**, 135-148 (2006).
12. G. Domes, M. Heinrichs, A. Michel, C. Berger, S. C. Herpertz, *Biological psychiatry* **61**, 731-733 (2007).
13. R. Wageman, J. R. Hackman, E. Lehman, *Journal of Applied Behavioral Science* **41**, 373-398 (2005).
14. J. P. Stokes, *Small Group Research* **14**, 163 (1983).
15. A. Edmondson, *Research on managing groups and teams: Context* **2**, 179-199 (1999).
16. A. W. Woolley, *Organization Science* **20**, 500-515 (2009).
17. A. Pentland, *Honest signals: How they shape our world* (Bradford Books, Cambridge, MA, 2008).