Clarifying the Structure of Collective Intelligence in Teams: A Meta-Analysis

VALERIE GRAF, University of Augsburg
HENNER GIMPEL, University of Augsburg
JORDAN B. BARLOW, University of St. Thomas

1. INTRODUCTION

General intelligence is a statistical measure that emerges from the correlations among how well individuals do a wide variety of tasks (Spearman 1904). By analogy, collective intelligence is defined as a measure of a group's capacity to perform a wide variety of tasks (Engel et al. 2014, Kim et al. 2016, Woolley et al. 2010). A seminal paper established the existence of collective intelligence in teams collaborating in a face-to-face context by measuring a single dominant factor that accounts for major variance in a team's task performance scores (Woolley et al. 2010). Though collective intelligence may be partially determined by general individual intelligence (Barlow 2017, Bates and Gupta 2017), other factors also play a role (Meslec et al. 2016, Chikersal et al. 2017, Kim et al. 2017).

The consistent existence of a collective intelligence construct would have implications for how organizational teams are assembled, trained, and even evaluated, as it suggests that a single set of competencies and processes underpin performance in all teams (Crede and Howardson 2017). Thus, there should be widespread interest in how it can be managed effectively in different organizational team collaboration contexts. However, replications in face-to-face, online, and hybrid contexts have all provided mixed results regarding the existence of a single collective intelligence factor (Engel et al. 2014, 2015, Bates and Gupta 2017, Barlow and Dennis 2016a, 2016b). The goal of this research is to re-analyze the existence and structure of collective intelligence across contexts of team collaboration, considering multiple data sets in a meta-analysis. Individual studies suffer from limits in scope and scale and, thus, fail to identify a potentially multifaceted structure of collective intelligence.

The existence, structure, and measurement of general individual intelligence has been one of the most researched and discussed topics in psychology over the last hundred years (Deary 2000). Traditional intelligence tests follow Spearman and are designed to measure cognitive ability in a single score (Nevid 2003). However, other theories argue that a single score cannot meaningfully reflect the full range of mental abilities and suggest models of multiple primary mental abilities (Thurstone and Thurstone 1941, Gardner 2008, Sternberg 1985). For example, a meta-analysis on more than 450 data sets conceptualizes individual intelligence as a higher-order factor model with general individual intelligence as the highest order factor, showing paths to several sub-factors such as cognitive speediness, visual perception, memory, learning, and further factors deeper down in the hierarchy (Carroll 1993). Just as Woolley et al. (2010) proposed a single factor of collective intelligence based on Spearman's work, we propose, like later individual intelligence researchers, that the structure of collective intelligence is more complex than a single factor.

2. METHODS AND RESULTS

We identified an initial set of scientific papers that used quantitative analysis of collective intelligence and included correlation coefficients (r or ρ) between employed team task types. Studies missing variable covariates and qualitative articles were excluded. Our first meta-analytic sample included 745 teams in 13 studies from 6 scientific papers: Woolley et al. 2010 (2 studies: 40, 107 teams); Engel et al. 2014 (2 studies: 32, 36 teams); Engel et al. 2015 (3 studies: 68, 25, 116 teams); Barlow & Dennis 2016a

(2 studies: 64, 65 teams); Barlow & Dennis 2016b (1 study: 86 teams); Bates & Gupta 2017 (3 studies: 26, 40, 40 teams). As work-in-progress, we have now performed a more comprehensive literature review and are in process of updating our meta-analysis to include additional studies that have been published more recently. We report here our initial results. The primary studies subject to this analysis used different tasks for measuring identical task types defined by McGrath (1984). In total, the identified studies in our initial set referenced 5 out of 8 task types: generating ideas (Type 2), solving problems with correct answers (Type 3), deciding issues with no right answer (Type 4), resolving conflicts of interests (Type 6), and executing performance tasks (Type 8).

Meta-analytic structural equation modeling (MASEM) for complex datasets (Wilson et al. 2016) served to synthesize respective correlation coefficients to a single pooled correlation matrix. Hence, the correlation matrices of the studies were synthesized to a pooled 5x5 correlation matrix referencing the employed task types. We used a three-level multivariate mixed-effects weighted meta-regression model to account for the respective complexities and derive a pooled correlation matrix. This model accounts for statistical dependencies associated with the clustering of units within levels, including those resulting from the use of multiple tasks within the same studies (Wilson et al. 2016). As Equation (1) illustrates, the dependent variable r_{ik} represents the observed correlation coefficients $i = 1 - 128$ from study $k = 1 - 13$. Each cell of the future pooled correlation matrix is represented with a unique independent dummy variable ($Cell_{1ik}$, …, $Cell_{10ik}$), which take a value of 1 if coefficient i from study k is assigned to that cell and a value of 0 otherwise. This serves to assign each effect size stated in the primary correlation matrices to its "right position" in the pooled matrix.

$$r_{ik} = \beta_1 Cell_{1ik} + \beta_2 Cell_{2ik} + \ldots + \beta_{10} Cell_{10ik} + \nu_{0k} + \eta_{ik} + \varepsilon_{ik} \quad (1)$$

The use of a no-intercept model permits interpreting the respective regression coefficients as pooled correlation coefficients. Furthermore, variables η_{ik} and ν_{0k} represent Level 2 and Level 3 random effects for identified studies and are assumed to be normally distributed with a mean of 0, variance $\omega > 0$, and $\tau > 0$. While Level 2 random effects capture random effects of all coefficients in the pooled matrix, Level 3 random effects capture random effects of all cells in the pooled matrix. The estimation error ε_{ik} is also assumed to be normally distributed with a mean of 0 and variance of υ_{ik}. The conditional sampling covariance between observed correlations from the same study is approximated by the unconditional Level 2 random effects. It is assumed that the errors at different levels are uncorrelated. Finally, as correlational effect sizes based on larger samples are more precise, they were weighted more heavily by applying inverse sample size weighting. As procedural and methodological variation in studies might obscure or even distort the underlying relationships of interest, respective variations must be smoothed out (Wilson et al. 2016). A moderator was introduced, approximating a situation where potential study-level effects originating from different collaboration contexts (face-to-face, online, hybrid) are smoothed out. Another three-level mixed effects meta-regression model (Equation (2)) served to predict correlation coefficients expected with selected values of the respective moderator X_{0k}.

$$r_{ik} = \beta_{00} + \beta_1 X_{0k} + \nu_{0k} + \eta_{ik} + \varepsilon_{ik} \quad (2)$$

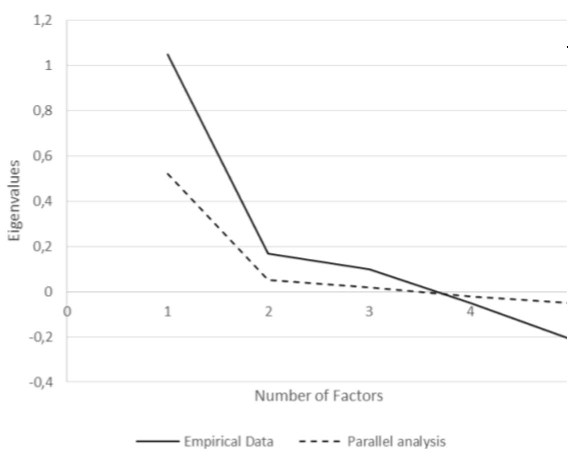
It is assumed that the regression coefficient for a moderator is the same across all cells of the matrix, which is a reasonable assumption in this case, as correlations are consistently larger in established experiments when teams collaborate in face-to-face contexts (Woolley et al. 2010) compared to CMC contexts (Barlow and Dennis 2016b). The random effects (ν_{0k}, η_{ik}) represent additional heterogeneity associated with experiment-level differences and differences between correlations due to size. Relationships represented by the coefficient β_1, reflect the influence of different experiment-level contexts that we wanted to adjust for consistency. The pooled correlation matrix resulting from the multi-level mixed effects meta-regression is displayed in Table 1.

	1	2	3	4
1 Generating Ideas				
2 Solving problems with correct answers	-0.17			
3 Deciding issues with no right answer	0.01	-0.02		
4 Resolving conflicts of interest	0.07	0.07	-0.09	
5 Executing performance tasks	0.17	0.13	-0.02	-0.05

Table 1. Pooled correlations among team task types, n=745 teams, aggregated via multi-level mixed effects meta-regression

The prevailing theory suggests the existence of a single dominant factor (Barlow and Dennis 2016a, Engel et al. 2014, 2015, Woolley et al. 2010). We first used a confirmatory factor analysis (CFA) to assess the fit of this theory with empirical evidence, specifically the newly synthesized correlation matrix. However, the CFA found a bad fit for the single-factor solution [$\chi^2 = 40.79$, P = 0.00 (>0.05 would indicate good fit; Jöreskog and Sörbom 1989), df = 5; χ^2/df = 8.2 (between 1-2 would indicate good fit; Hair et al. 1995, 1998); CFI = 0.73 (>0.90 would indicate good fit; Bentler 1990); TLI = 0.45 (>0.90 would indicate good fit; Tucker and Lewis 1973); RMSEA = 0.10 (<0.05 or <0.08 would indicate a good fit; Browne and Cudeck 1993)], suggesting that the theory is not compatible with the data.

We next applied principal axis factoring (PAF), which, like the CFA results, did not support the theory that a single collective intelligence factor exists (Figure 1). With exploratory factor analysis, the first factor can be declared a single intelligence factor if it has an eigenvalue >1.38 and accounts for 30−50% of variance (Woolley et al. 2010). Our factor analysis yielded three factors instead of a single dominant one. The first factor showed an eigenvalue of 1.05, which was lower than the threshold of 1.38 and only accounted for 17% of variance. The second factor did not explain significantly less than the first one (15%). Further, the average inter-item correlation for teams' scores on different task types was low (r=0.001). Parallel analysis (Horn 1965) clearly suggested that a three-factor model is the best fit for the synthesized empirical data (Figure 1).



Factors:	1	2	3
1 Generating Ideas	**.89**	.04	.15
2 Solving problems with correct answers	-.22	.10	.15
3 Deciding issues with no right answer	.02	-.09	-.02
4 Resolving conflicts of interest	.05	**.99**	-.07
5 Executing performance tasks	.03	.01	**.90**

Fig. 1. Left: Parallel analysis yields 3 factors similarly accounting for the variance in the teams' performance scores on different task types. Right: Factor loadings of team tasks types after Varimax rotation.

The cumulative empirical evidence does not support the existence of a single inherent collective intelligence factor that transcends team collaboration contexts and a wide variety of cognitive tasks. Instead, we found team performance to be structured by multiple factors—specifically, a three-factor structure of collective intelligence that consists of (1) idea generation or creativity, (2) conflict resolution, and (3) execution of tasks. These results provide more insight into the complexity of collective intelligence and help shed light on the mixed findings in prior literature.

REFERENCES

Barlow, J. B. 2017. Collective Intelligence and its relationship to collective individual intelligence. *Collective Intelligence Conference*.

Barlow, J. B., A. R. Dennis. 2016a. Group coordination structuring: A process to facilitate collective intelligence in virtual group work. *Collective Intelligence Conference*.

Barlow, J. B., A. R. Dennis. 2016b. Not as smart as we think: A study of collective intelligence in virtual groups. *Journal of Management Information Systems* **33**(3) 684–712.

Bates, T. C., S. Gupta. 2017. Smart groups of smart people: Evidence for IQ as the origin of collective intelligence in the performance of human groups. *Intelligence* **60** 46–56.

Bentler, P. M. 1990. Comparative fit indexes in structural models. *Psychological Bulletin* **107**(2) 238–246.

Browne, M. W., R. Cudeck. 1993. Alternative ways of assessing model fit. K. A. Bollen, J. S. Long, eds. *Testing structural equation models*. SAGE Publications, Newbury Park, CA, 136–162.

Carroll, J. B. 1993. *Human cognitive abilities. A survey of factor-analytic studies.* Cambridge University Press, Cambridge, NY.

Chikersal, P., M. Tomprou, Y.J. Kim, A.W. Woolley, L. Dabbish. 2017. Deep structures of collaboration: Physiological correlates of collective intelligence and group satisfaction. *ACM Conference on Computer Supported Collaborative Work and Social Media (CSCW).*

Crede, M., G. N. Howardson. 2017. The structure of group task performance: A second look at the evidence for the "collective intelligence" construct. *Accepted for Publication at: Journal of Applied Psychology*.

Deary, I. J. 2000. *Looking down on human intelligence. From psychometrics to the brain.* Oxford University Press, Oxford, Great Britain.

Engel, D., A. W. Woolley, I. Aggarwal, C. F. Chabris, M. Takahashi, K. Nemoto, C. Kaiser, Y. J. Kim, T. W. Malone. 2015. Collective intelligence in computer-mediated collaboration emerges in different contexts and cultures. *Conference for Human-Computer Interaction (CHI).*

Engel, D., A. W. Woolley, L. X. Jing, C. F. Chabris, T. W. Malone. 2014. Reading the mind in the eyes or reading between the lines? Theory of mind predicts collective intelligence equally well online and face-to-face. *PloS One* **9**(12).

Gardner, H. E. 2008. *Multiple intelligences. New horizons in theory and practice.* Basic books, New York, NY.

Hair, J. F., R. E. Anderson, R. L. Tatham, W. C. Black. 1995. *Multivariate data analysis with Readings.* Prentice Hall, Englewood Cliffs, NJ.

Hair, J. F., R. E. Anderson, R. L. Tatham, W. C. Black. 1998. *Multivariate data analysis with Readings.* Prentice Hall, Englewood Cliffs, NJ.

Horn, J. L. 1965. A rationale and test for the number of factors in factor analysis. *Psychometrika* **30**(2) 179–185.

Jöreskog, K. G., D. Sörbom. 1989. *LISREL 7. A guide to the program and applications.* SPSS Inc., Chicago, IL.

Kim, Y. J., I. Aggarwal, A. W. Woolley. 2016. How collective intelligence impacts team performance: Exploring transactive memory systems as a mechanim. *Collective Intelligence Conference*.

Kim, Y.J., D. Engel, A.W. Woolley, J.Y. Lin, N. McArthur, T.W. Malone. 2017. What makes a strong team? Using collective intelligence to predict team performance in League of Legends. *ACM Conference on Computer Supported Collaborative Work and Social Media (CSCW).*

Meslec, N., I. Aggarwal, P.L. Curseu. 2016. The insensitive ruins it all: Compositional and compilational influences of social sensitivity on collective intelligence in groups. *Frontiers in Psychology* **7**(676).

McGrath, J. E. 1984. *Groups: Interaction and Performance.* Prentice-Hall, Englewood Cliffs, NJ.

Nevid, J. 2003. *Psychology. Concepts and applications.* Wadsworth, Belmont, CA.

Spearman, C. 1904. General Intelligence, objectively determined and measured. *The American Journal of Psychology* **15**(2) 201–292.

Sternberg, R. J. 1985. *Beyond IQ. A triarchic theory of human intelligence.* Cambridge University Press, Cambridge, NY.

Thurstone, L. L., T. G. Thurstone. 1941. Factorial studies of intelligence. *Psychometric Monographs* **2**.

Tucker, L. R., C. Lewis. 1973. A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* **38**(1) 1–10.

Wilson, S. J., J. R. Polanin, M. W. Lipsey. 2016. Fitting meta-analytic structural equation models with complex datasets. *Research Synthesis Methods* **7**(2) 121–139.

Woolley, A. W., C. F. Chabris, A. Pentland, N. Hashmi, T. W. Malone. 2010. Evidence for a collective intelligence factor in the performance of human groups. *Science* **330**(6004) 686–688.