

Graf-Drasch, Valerie; Gimpel, Henner; Barlow, Jordan B.; Dennis, Alan R.

Article — Published Version

Task structure as a boundary condition for collective intelligence

Personnel Psychology

Provided in Cooperation with:

John Wiley & Sons

Suggested Citation: Graf-Drasch, Valerie; Gimpel, Henner; Barlow, Jordan B.; Dennis, Alan R. (2021) : Task structure as a boundary condition for collective intelligence, Personnel Psychology, ISSN 1744-6570, Wiley, Hoboken, NJ, Vol. 75, Iss. 3, pp. 739-761, <https://doi.org/10.1111/peps.12489>

This Version is available at:

<https://hdl.handle.net/10419/265026>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<http://creativecommons.org/licenses/by/4.0/>

Task structure as a boundary condition for collective intelligence

Valerie Graf-Drasch¹  | Henner Gimpel¹ | Jordan B. Barlow² | Alan R. Dennis³

¹ University of Hohenheim, and FIM Research Center, Project Group Business & Information Systems Engineering of the Fraunhofer FIT, Germany

² University of St. Thomas

³ Kelley School of Business, Indiana University

Correspondence

Valerie Graf-Drasch, University of Hohenheim, Schloß Hohenheim 1, 70599 Stuttgart, Germany.
Email: valerie.graf-drasch@uni-hohenheim.de

This research was supported in part by the German Research Foundation (DFG; Grant Number 343128888).

Abstract

In 2010, a new research stream began on collective intelligence (CI), defined as a group's general ability to perform consistently well across a wide variety of tasks. Subsequent empirical evidence presents a mixed picture. Some studies have found groups to exhibit CI while others have not. To resolve these disparate results, we compare 21 experimental studies to understand what influences whether groups exhibit CI. We find that task structure is a boundary condition for CI in that groups exhibit CI across well-structured tasks but not across ill-structured tasks. For ill-structured tasks, CI has a more nuanced set of multiple factors that may be interpreted as different facets of CI. This research extends our understanding of CI by suggesting that the original definition of CI was too all-encompassing. CI should be reconceptualized as a multi-dimensional phenomenon, similar to research on individual intelligence. We highlight avenues for future research to continue to move CI research forward, particularly regarding ill-structured tasks.

KEYWORDS

cognitive psychology, collective intelligence, group performance

1 | INTRODUCTION

Intelligence is an individual trait, yet groups may also exhibit intelligence (Malone & Bernstein, 2015). The concept of Collective Intelligence (CI) was initially coined by the entomologist Wheeler (1911), who observed ants cooperating in the colony so closely as to become indistinguishable from a single organism exhibiting intelligent behavior. Other researchers have studied this phenomenon using the terms “collective intelligence” (Malone & Bernstein, 2015; Sulis, 1997, 2009), “group intelligence” (Williams & Sternberg, 1988), “wisdom of crowds” (Surowiecki, 2004), “vox populi” (Galton, 1907), “collective mind” (Weick & Roberts, 1993), or “organization mind” (Sandelands & Stablein, 1987).

Over the last decade, Woolley et al (2010) popularized a new interpretation of CI, defined as a (small) group’s general ability to perform a wide variety of tasks (Woolley et al., 2010). Thus, groups high in CI would perform consistently well across a set of tasks, and groups low in CI would perform consistently poorly. This definition of CI has roots in the theory of individual intelligence. Specifically, Woolley et al. (2010) adopted theoretical and measurement arguments from general intelligence as documented by Spearman (1904). Doing so, Woolley et al. (2010, p. 867) note that “the empirical fact of general cognitive ability as first demonstrated by Spearman is now, arguably, the most replicated result in all of psychology.” Yet, the CI factor results have been much more difficult to replicate.

The essence of individual or CI is the *ability to perform*. Intelligence assessments focus on measuring performance (Spearman, 1904; Woolley et al., 2010). The seminal work by Woolley et al. (2010) measured CI with an exploratory factor analysis (EFA), which found a single general factor that accounted for substantial variance in a group’s performance scores across a variety of tasks. Other research followed this approach, with many finding that groups exhibit a single general factor (Engel et al., 2015, 2014; Gupta et al., 2019; Kim et al., 2017; Meslec et al., 2016; Ostrowski et al., 2019; Riedl et al., 2021). Others have not found groups to exhibit CI (Barlow & Dennis, 2016a, 2016b; Bates & Gupta, 2017; Gimpel & Graf-Drasch, 2020; Rowe, 2019). At this point, it is unclear whether CI exists as originally defined or whether the definition needs to be refined. In 2017, Credé et al. (2017) concluded that “if a factor exhibits a strong relationship with scores on only some group tasks but not with other group tasks, then we cannot reasonably infer that performance on the full set of group tasks is characterized by a general [CI] factor” (p. 1485).

Given this prior research, we pursue two aims. First, we aim to understand what boundary conditions influence whether groups exhibit CI by examining existing empirical research. CI is a group’s ability to perform consistently well across tasks (Woolley et al., 2010), so one factor that is likely to play a critical role is the structure of those tasks. Credé et al. (2017) noted that even in studies that find a single CI factor, this factor is related to some tasks and not to others, indicating that the nature of the task is an important element in whether groups exhibit CI or not. Hence a group’s ability to perform well across a variety of tasks may depend on the tasks’ fundamental structure—well or ill-structured. A task is defined as well-structured when it has a predetermined, verifiably correct solution. A task is defined as ill-structured when it lacks predetermined or demonstrably correct solutions (Schraw et al., 1995; Wildman et al., 2012). Ill-structured tasks require different coordination and cognitive processes than well-structured tasks. Many well-structured tasks can be tackled similarly, while ill-structured tasks require unique approaches (Newell & Simon, 1972; Schraw et al., 1995; Simon, 1973).

Second, if groups do not exhibit CI as defined by Woolley et al. (2010) for either well-structured or ill-structured tasks, then it is important to understand what pattern of factors groups do exhibit. If the factor analysis does not produce a single general factor that Woolley et al. (2010) define as the indicator of CI, what is the nature of performance that groups exhibit? More formally, we propose the following two research questions:

Research question 1 (RQ1): *Does task structure (i.e., well-structured or ill-structured) influence whether groups exhibit CI?*

Research question 2 (RQ2): *If groups do not exhibit CI for well-structured or for ill-structured tasks, what pattern of factors do they exhibit?*

These research questions are sequential in the sense that the second one builds on the first. If the answer to RQ1 is no, RQ2 is meaningless. If RQ1 is answered with a yes, then RQ2 sheds more light on the nature of the difference. To answer the first question, we examine 21 studies and analyze the underlying structure of the tasks they use. To answer the second research question, we conduct a meta-analysis of prior studies.

Our investigation provides three contributions to the literature. First, we contribute by clarifying the boundary conditions for CI as theorized and measured by Woolley et al. (2010). We examine a variety of boundary conditions and find that a significant boundary condition is task structure in that groups exhibit CI across well-structured tasks but not across ill-structured tasks. Our examination also found six other factors (group size, cultural contexts, collocation, collaboration medium, payment scheme, and group context) to not be boundary conditions for CI in our sample.

Second, in finding the boundary condition of task structure, we offer new theoretical insight and potential refinement of the CI definition and its measurement. The original definition of CI as a (small) group's general ability to perform a wide variety of tasks (Woolley et al., 2010) did not constrain CI to only well-structured tasks, yet the original study by Woolley et al. (2010) used a set of primarily well-structured tasks. This approach was adopted by most researchers, who found similar results for a general CI factor. Other researchers found different results when they examined CI using a mix of well-structured and ill-structured tasks. Our meta-analysis finds a single general CI factor for well-structured tasks but not for ill-structured tasks. One important next step for CI research is thus to align the theoretical definition of CI with its measurement so that different researchers share a common understanding. To help advance research on CI, we propose redefining CI as a group's ability to perform consistently across a variety of well-structured tasks and continuing to measure it as done by Woolley et al. (2010).

In refining the definition and measurement of CI to well-structured tasks, our third contribution is in opening up future research to investigate CI in the context of ill-structured tasks. Our meta-analysis found that for ill-structured tasks, CI has a more nuanced set of multiple factors that may be interpreted as different facets of CIs. Future research may take the domain of individual intelligence as a starting point. A single factor concept of individual intelligence is challenged by theories suggesting individuals have multiple distinct intelligences such as creative, analytic, practical, linguistic, logical-mathematical individual intelligence (e.g., Gardner, 2008; Sternberg, 1985). Building from this research, we suggest that groups may also have multiple distinct CI.

2 | THEORETICAL BACKGROUND

2.1 | Collective intelligence

The concept of CI originates from the concept of general intelligence, first documented by Charles Spearman (Spearman, 1904; Woolley et al., 2010). Spearman defined that intelligence reflects an invisible "mental energy" which, like gravity, underlies and constrains individual performance on a variety of cognitive tasks (Demetriou, 2002; Spearman, 1904). Individuals who do well (or poorly) on one set of tasks are likely to do well (or poorly) on other tasks. To measure individual intelligence, Spearman performed a factor analysis and noted that a single general factor (*g*) explained a large part of the variance in the cognitive performance of individuals across a diverse set of tasks. Importantly, intelligence is not the same as *g*. Intelligence is a broad category referring to cognitive mental processes like stimulus apprehension, perception, attention, discrimination, generalization, learning, memory, inference, reasoning, and language (Jensen, 2002). "The *g* factor is something else [...] reflecting *individual differences* in performance on tests or tasks that involve any one or more of the kinds of processes just referred to as *intelligence*" (Jensen, 2002, p. 40 italics in the original).

Woolley et al. (2010) transferred the theory and measurement around *g* from research on individual intelligence to the domain of groups. "By analogy with individual intelligence, we define a group's collective intelligence (*c*) as the general ability of the group to perform a wide variety of tasks" (Woolley et al., 2010, p. 687). Groups that do well (or poorly) on one set of tasks are likely to do well (or poorly) on other tasks (Woolley et al., 2010). In analogy with Spear-

man, Woolley et al. (2010) measured CI by conducting a factor analysis and found that a single general CI factor (c) accounted for substantial variance in groups' performance across a set of tasks. They defined this factor as a measure of CI.

The concept of CI is related to other group-level constructs such as *group performance*, *group effectiveness*, or *group potency* but is fundamentally different. Group performance is the outcome of one task, whereas CI is a group's ability to perform, so it is *measured* using group performance across a variety of tasks. In that, CI differs from *group effectiveness* (Alper et al., 2000; Barrick et al., 1998; LePine et al., 2008) because performance is just one of many dimensions required to measure group effectiveness, which also includes dimensions such as member attitudes and behavioral outcomes (S. G. Cohen & Bailey, 1997; Hackman & Morris, 1975). If a group exhibits CI, this may lead, in part, to group effectiveness. *Group potency* is the belief that a group can be effective (Campion et al., 1993; Hu & Liden, 2011; Shea & Guzzo, 1987). CI differs from group potency because group potency is a belief, whereas CI is an actual ability.

Research has both supported and challenged the argument that groups exhibit CI. The findings of Woolley et al. (2010) have been successfully replicated in face-to-face as well as in computer-mediated communication contexts, showing that CI transcends media (Engel et al., 2014, 2015; Meslec et al., 2016; Woolley et al., 2017). CI may be partially influenced by general individual intelligence (Barlow, 2017; Bates & Gupta, 2017), but other factors also play a role (Chikersal et al., 2017; Kim et al., 2017; Meslec et al., 2016). Other studies have challenged these results in that they have failed to find empirical evidence of CI as defined by the seminal Woolley et al. (2010) study (Barlow & Dennis, 2016a, 2016b; Bates & Gupta, 2017).

One potential reason for the mixed empirical results centers on the *task* because CI is defined as a group's general ability to perform consistently well across a wide variety of *tasks* (Woolley et al., 2010). Woolley and colleagues used short problems with demonstrably correct solutions that do not require much coordination within groups (also called "Eureka tasks," e.g., Sudoku puzzles or word searches; Barlow & Dennis, 2016b), whereas other studies used tasks that were less structured and/or required more coordination.

2.2 | Task types and task structures

In order to reconcile the mixed results for CI, it becomes important to understand the variety of tasks that have been utilized in CI research. To examine this variety, prior empirical CI studies have used different types of tasks proposed by McGrath's (1984) task circumplex in their experiments. See the first three columns in Table 1.

However, research on cognitive psychology argues that a group's performance often depends less on task type than on a task's fundamental structure (either well-structured or ill-structured) requiring different cognitive processes (Newell & Simon, 1972; Schraw et al., 1995; Voss & Post, 1988). This is not to say that task types defined by McGrath (1984) are unimportant. Rather, they are a different level of analysis. Several different task types can have similar task properties (Fischhoff & Broomell, 2020) and can thus either be classified as well-structured or ill-structured (Schraw et al., 1995). For example, one important property of any task is the presence or absence of a verifiably correct answer. Well-structured tasks are characterized by verifiably correct answer(s), while ill-structured tasks are characterized by a lack of predetermined or demonstrably correct solutions (Jonassen, 1997; Schraw et al., 1995). Thus, different ill-structured tasks often require different cognitive processes such as increased coordination, whereas different well-structured tasks often share common elements, such that similar cognitive processes can be used (Jonassen, 1997; Schraw et al., 1995).

Research often uses two criteria to identify a well-structured task, namely (1) one or more verifiably correct solutions (i.e., solution demonstrability) and/or the presence of several decision options and (2) a guaranteed procedure available to reach that solution (Gick, 1986; Jonassen, 1997; Schraw et al., 1995; Simon, 1973; Sinott, 1989). Examples of well-structured tasks are reasoning and intellectual construction, logical and precision tasks, solving problems with

TABLE 1 Task types defined by McGrath (1984) and classification by task structure

Quadrant and type		Description	Classification as well-structured or ill-structured task based on the operationalization in the primary studies	Exemplary task operationalization
QUADRANT I		GENERATE		
Type 1:	Planning task	Generating plans	Ill	Party planning
Type 2:	Creativity task	Generating ideas	Ill	Brainstorming
QUADRANT II		CHOOSE		
Type 3:	Intellective task	Solving problems with a correct answer	Tendentially well	Matrix reasoning, Sudoku, estimation, missing letters/unscramble words
Type 4:	Decision-making task	Dealing with tasks for which the preferred or agreed-upon answer is the correct one	Tendentially ill	Judgment, estimation problems
QUADRANT III		NEGOTIATE		
Type 5:	Cognitive conflict task	Resolving conflicts of viewpoint (not of interest)	Ill	Moral reasoning
Type 6:	Mixed-motive task	Resolving conflicts of motive interest	Ill	Shopping trip
QUADRANT IV		EXECUTE		
Type 7:	Contests/battles	Resolving conflicts of power, competing for victory	Well	Reproducing art, memory
Type 8:	Performances	Psychomotor tasks performed against objective or absolute standards of excellence	Well	Group typing

single correct answers, clerical tasks (e.g., checking a database against a paper copy), or discussion tasks (i.e., group discusses a question; Wildman et al., 2012).

By contrast, ill-structured tasks lack definition in some aspect(s) (Simon, 1973). They have multiple solutions (or even no solution at all), there is no guaranteed procedure to reach a solution, and the information needed to solve the task is not entirely contained in the problem (Newell & Simon, 1972; Simon, 1973). Ill-structured tasks often require generating a new, unique solution, a solution that is unlike solutions to other tasks (Newell & Simon, 1972; Simon, 1973). Because problem-solvers must select a good solution among many, they often need to generate a viable, defensible, and cogent argument to support the solution (Ge & Land, 2003; Jonassen, 1997; Voss & Post, 1988). Examples of ill-structured tasks are creativity, planning, and aesthetic design (Wildman et al., 2012).

Further properties differentiating well-structured and ill-structured tasks are coordination efforts and cognitive mechanisms required. Coordination research suggests people perform *production* and *coordination* when put into

groups (Malone & Crowston, 1990, 1994). *Production* is directly associated with goal achievement, is performed individually, does not involve interdependencies among actors, and is likely related to individual intelligence. Conversely, *coordination* refers to managing interdependencies between actors. Coordination can be divided into four different types—communication, perception of common objects, group decision making, and pure coordination (Malone & Crowston, 1990).

Group tasks usually have both production and coordination components, with some tasks being particularly high in coordination, while others are mainly a function of individual production activities and thus have little coordination (Kittur et al., 2009). Because much effort is needed to understand an ill-structured task (Simon, 1973), ill-structured tasks are often particularly high in coordination, with one major exception. Creativity tasks, such as brainstorming, are ill-defined tasks requiring little coordination because group performance is the sum of individual work. While some coordination may be beneficial, groups can perform brainstorming with no coordination and suffer no loss in performance (Diehl & Stroebe, 1987). Typical coordination activities of ill-structured tasks refer to communicating, assessing common objects, reaching consensus, allocating resources, and assigning activities (Malone & Crowston, 1990). Some of these coordination activities may also occur in well-structured tasks, but not to this extent. Adding structure to ill-structured tasks can make them well-structured (Simon, 1973).

Finally, well-structured and ill-structured tasks also require different cognitive problem-solving mechanisms (Newell & Simon, 1972; Schraw et al., 1995; Simon, 1973). In this regard, Kitchener (1983) developed a model comprising three levels of cognitive processing skills. Well-structured tasks require Kitchener's first two levels of cognitive processing, namely (1) inferential rules and strategies, as well as (2) processes such as metacognition to select and monitor level 1 skills. Conversely, ill-structured tasks also require level 3 skills, including epistemic monitoring to examine the legitimacy of solutions (Kitchener, 1983; Schraw et al., 1995).

Therefore, well-structured and ill-structured tasks require different problem-solving processes, skills, and characteristics. Specifically, solving well-structured tasks requires the three basic procedural steps of understanding the task, searching for a solution, and implementing a solution (Ge & Land, 2003; Gick, 1986). Groups engaged in well-structured tasks may profit from clear and designated structures reflected in pre-defined workflows and a designated team leader (Wildman et al., 2012). By contrast, solving ill-structured tasks is more complex and requires different steps, such as applying the selected solution against alternatives (Ge & Land, 2003; Sinott, 1989; Voss & Post, 1988). Thus, for ill-structured tasks, team members' abilities to understand and justify actions are paramount (Ge & Land, 2003; Jonassen, 1997; Kitchener & King, 1981). Groups working on ill-structured tasks may profit from a working environment encouraging creativity and openness, facilitation of multiple diverse perspectives, as well as distributed leadership or functional role structures (Wildman et al., 2012).

In summary, the current state of research on CI, along with theory and research on the structure of group tasks, leads us to two research questions. Current CI research presents conflicting conclusions as to whether groups exhibit CI. Our first goal is to understand what boundary conditions influence whether groups exhibit CI. Second, research suggests that well-structured and ill-structured tasks differ in their core properties, such as the presence of a verifiable correct answer, coordination efforts, and cognitive mechanisms required. Thus, RQ1 is: *Does task structure (i.e., well-structured or ill-structured) influence whether groups exhibit CI?* If it does, we follow up with RQ2: *If groups do not exhibit CI for well-structured or for ill-structured tasks, what pattern of factors do they exhibit?*

3 | METHOD

3.1 | Literature search and inclusion criteria

To identify relevant research papers, we performed a comprehensive literature review consisting of three different search strategies. First, we used the Web of Science and PsycINFO databases with the search term *collective*

intelligence (in quotation) from 2010 to 2020 (until January 5). We also examined the references in these papers to find additional studies. Second, we followed the “Discourse Approach” of Larsen et al. (2019) by examining all papers that cited the foundational article of Woolley et al. (2010). We used Microsoft Academic Search (MAS; academic.research.microsoft.com) and Google Scholar, both of which include “gray literature” (e.g., books, book chapters, monographs, non-peer-reviewed conference proceedings, and unpublished articles like the ones from SSRN). Third, to ensure we did not miss research not publicly available, we requested papers under revision or not (yet) published articles from researchers in the CI community who attended the ACM Collective Intelligence Conference in 2019 and 2020.

The resulting literature set contains articles that contribute to developing the CI construct and articles that use CI for other research purposes. The objective of this paper is to understand what conditions enable or inhibit groups from exhibiting CI. Therefore, we selected all papers that empirically tested whether or not groups exhibited CI. We excluded studies that measured CI as a construct but did not provide empirical detail on construct validity (e.g., Chikersal et al., 2017). We also excluded studies that contained data from previous studies already considered in our set (e.g., Aggarwal et al., 2019; Riedl et al., 2021; Woolley et al., 2017, and the U.S. sample of Engel et al., 2015).

We identified a total of 12 research papers with 21 separate studies for analysis. Table 2 lists the studies, the nature of the tasks, and other factors.

3.2 | Coding procedure

First, one author coded whether or not the researchers concluded that groups exhibited CI in each of the 21 studies. This information was objectively available as each study reported the results of an EFA, which either resulted in a single general factor structure representing the CI factor or not (e.g., multi-factor structure instead).

Second, the coding of the task as well-structured or ill-structured was done independently by two researchers (one author and one postdoctoral researcher). The 21 studies explicitly reported task types as defined by McGrath (1984) and descriptions of each task operationalization but no categorization into well-structured or ill-structured tasks. Therefore, the coders looked at the task descriptions indicating how the McGrath (1984) task types were operationalized in the 21 studies to make a determination to code as well-structured or ill-structured. The determination was made on the basis of the criteria of well-structured and ill-structured tasks summarized in Table 3. While six of the eight McGrath (1984) task types were consistently operationalized as well-structured or ill-structured tasks across studies, two McGrath (1984) task types varied in their operationalizations across studies (i.e., task type 3 “intellective tasks” and task type 4 “decision-making tasks”). In some studies, these tasks were operationalized as well-structured, while other studies operationalized them as ill-structured. Detailed descriptions of all task operationalizations are provided in Online Appendix A. A short version is captured in the last two columns of Table 1. Overall, the two coders had a Cohen’s Kappa of .86, an “almost perfect” agreement level (Landis & Koch, 1977).

Lastly, for each study, we counted the number of tasks we determined as well-structured and divided it by the total number of tasks used in this study (i.e., well-structured and ill-structured tasks). This procedure yielded a variable termed “percentage of well-structured tasks” for each of the 21 studies, ranging from 0% to 100%. We will use this variable in the following analyses.

4 | RQ1: ANALYSIS AND RESULTS

First, we conducted three separate analyses on the 21 studies listed in Table 2 to investigate if task structure influences whether studies find that groups exhibit CI. Taken together, these three analyses point to task structure as a boundary condition for CI.

TABLE 2 Studies resulting from literature research

ID	Research paper	Sample size (number of groups)	Group size (range)	Group size (midpoint of range)	Number of well-structured tasks	Percent-age of well-structured tasks	Cultural context	Collocation	Collaboration medium	Payment scheme	Group context	Exhibition of CI	Study was part of RQ1 analyses	Study was part of RQ2 analyses
1	Woolley et al. (2010)	40	3.0	3.0	3, 2	60%	North America	Same place	Face-to-face	Equal payment	Ad-hoc	Yes	✓	✓
2		107	2-5	3.5	6, 4	60%	North America	Same place	Face-to-face	Equal payment	Ad-hoc	Yes	✓	✓
3	Engel et al. (2014)	32	4.0	4.0	6, 2	75%	North America	Same place	Face-to-face	Equal payment	Ad-hoc	Yes	✓	✓
4		36	4.0	4.0	6, 2	75%	North America	Same place	Chat	Equal payment	Ad-hoc	Yes	✓	✓
5	Engel et al. (2015)	116	2-5	3.5	5, 2	71%	Europe	Same place	Various media	Equal payment	Intact	Yes	✓	✓
6		25	4.0	4.0	5, 1	83%	Asia	Same place	Chat	Equal payment	Ad-hoc	Yes	✓	✓
7	Barlow and Dennis (2016b)	86	3-5	4.0	0, 3	0%	North America	Same place	Chat	Equal payment	Ad-hoc	No	✓	✓
8	Barlow and Dennis (2016a)	64	3-5	4.0	2, 0	100%	North America	Same place	Chat	Equal payment	Ad-hoc	Yes	✓	✓
9		65	3-5	4.0	0, 2	0%	North America	Same place	Chat	Equal payment	Ad-hoc	No	✓	✓
10	Meslec et al. (2016)	30	3-6	4.5	2, 1	67%	Europe	Same place	Chat	Equal payment	Intact	Yes	✓	✓

(Continues)

TABLE 2 (Continued)

ID	Research paper	Sample size (number of groups)	Group size (range)	Group size (midpoint of range)	Number of well-structured tasks	Percent-age of well-structured tasks	Cultural context		Collaboration medium		Payment scheme	Group context	Exhibition of CI	Study was part of	
							Cultural context	Collocation	Face-to-face	Face-to-face				RQ1 analyses	RQ2 analyses
11	Bates and Gupta (2017)	26	2-4	3.0	1, 2	33%	Europe	Same place	Face-to-face	Equal payment	Ad-hoc	No	✓	✓	
12		40	3.0	3.0	2, 3	40%	Asia	Same place	Face-to-face	Equal payment	Ad-hoc	No	✓	✓	
13	Bates and Gupta (2017)	40	3.0	3.0	2, 3	40%	Europe	Same place	Face-to-face	Equal payment	Ad-hoc	No	✓	✓	
14	Kim et al. (2017)	248	5.0	5.0	5, 1	83%	North America	Dispersed	Chat	Equal payment	Intact	Yes	✓	✓	
15	Gimpel and Graf-Drasch (2020)	50	3.0	3.0	4, 4	50%	Europe	Same place	Face-to-face	Performance-based	Ad-hoc	Yes	✓	✓	
16		50	3.0	3.0	4, 4	50%	Europe	Same place	Video and audio	Performance-based	Ad-hoc	Yes	✓	✓	
17		50	3.0	3.0	4, 4	50%	Europe	Same place	Video and audio	Performance-based	Ad-hoc	No	✓	✓	
18		50	3.0	3.0	4, 4	50%	Europe	Same place	Chat	Performance-based	Ad-hoc	No	✓	✓	
19	Gupta et al. (2019)	136	3-4	3.5	6, 2	75%	North America	Dispersed	Chat	Equal payment	Ad-hoc	Yes	✓	✓	
20	Ostrowski et al. (2019)	99	3.0	3.0	4, 0	100%	North America*	Same place	Various media*	Equal payment*	Ad-hoc	Yes	✓	✓	
21	Rowe (2019)	29	2-5	3.5	2, 3	40%	Australia	Same place	Face-to-face	Equal payment	Ad-hoc	No	✓	✓	

Note. Table represents information provided in primary studies, communication with authors of primary studies, and our analysis (information based on our analyses is marked with an asterisk*). Also, see Online Appendix A for further study information.

TABLE 3 Coding spreadsheet for tasks as well-structured or ill-structured (as particularly defined by Newell & Simon, 1972; Simon, 1973; Wildman et al., 2012)

Category	Description	Exemplary tasks
Well-structured	The presence of a verifiably correct answer or solution (i.e., solution demonstrability) and/or presence of several decision options (e.g., yes/no problems, choosing between paths A, B, C) rather than generating a new, unique solution to a given problem	Logical, precision tasks
		Reasoning and intellectual construction tasks
		Solving problems with correct answers (e.g., math problems)
		Clerical tasks (e.g., checking a database against a paper copy and correcting errors)
Ill-Structured	Lack of predetermined or demonstrably correct solutions and thus, often innovative, and creative tasks	Brainstorming and generation of ideas (e.g., brainstorming the design of a new product, generating ideas for a marketing campaign)
		Intellectual and analytical tasks that require the generation, exploration, and verification of knowledge
		Generating plans

4.1 | Graphical analysis

We plotted the studies to see the relationship between the number of well-structured versus ill-structured tasks and whether studies conclude that groups exhibit CI (see Figure 1). In this figure, each check or “x” represents one study. This figure reveals a clear result as *every study* that had more than 50% well-structured tasks found groups to exhibit CI, whereas *no study* that had fewer than 50% well-structured tasks did. The studies reported by Gimpel and Graf-Drasch (2020) included exactly 50% of well-structured tasks. Groups exhibited CI in half of the studies.

4.2 | Contingency analysis

We investigated whether there is a statistically significant and meaningful association between the percentage of well-structured tasks used in a study and whether groups exhibit CI. Groups exhibiting CI or not is a binary outcome. The percentage of well-structured tasks is a continuous variable ranging from 0% to 100%. However, given the limited set of empirical studies we have at hand, the actual percentage of well-structured tasks in our sample has only ten discrete values (e.g., 0% in the study by Barlow & Dennis, 2016b; 33% in one study by Bates & Gupta, 2017; and so on). Thus, in this contingency analysis, we treated the percentage of well-structured tasks as taking on discrete values and building a two (groups exhibiting CI or not) by ten (percentages of well-structured tasks) contingency table. Data in each cell present the number of observations (i.e., number of studies). In the subsequent multiple regression analysis, we treated the percentage of well-structured tasks as a continuous variable.

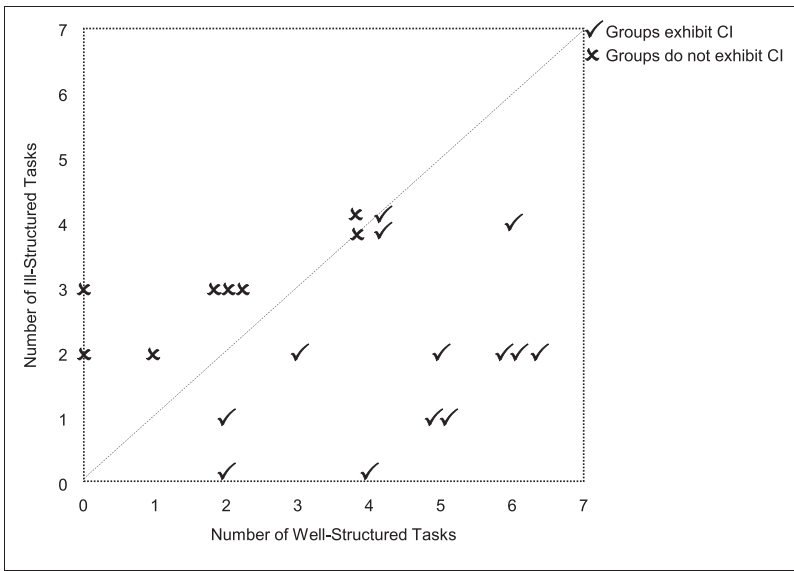
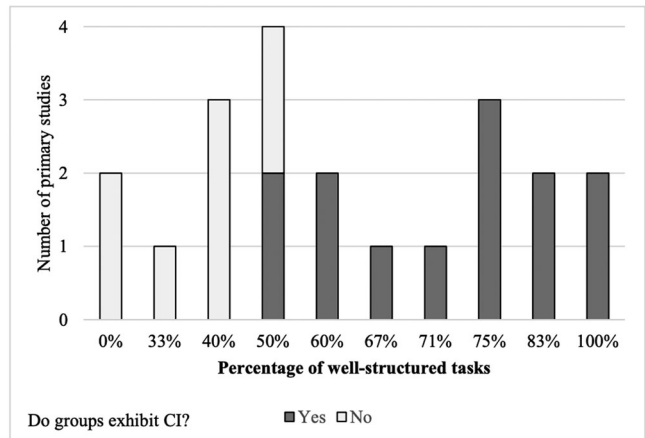


FIGURE 1 Scatterplot of when groups exhibit CI

FIGURE 2 Stacked bar chart depicting association between percentage of well-structured tasks and when groups exhibit CI



Adj. C: .94, p-value: .011, effect size: .89 large

We used Pearson’s contingency coefficient (C) to assess differences in the contingency table statistically. C is a measure of association between two categorical variables and uses chi-square to compare data in a contingency table (Sheskin, 2011). We used the adjusted contingency coefficient of Ott et al. (1992) that enables comparisons across differently sized contingency tables. We evaluated the significance of association (i.e., C) via the chi-square test (Agresti & Kateri, 2011; Sheskin, 2011) with Monte Carlo simulation, as suggested by Hope (1968). We evaluate the meaningfulness by following J. Cohen (1988), who translated C in an effect size (w), which he terms large if $w \geq .5$, medium if $w \geq .3$, and weak if $w \geq .1$. We used the statistical software R to conduct the analyses.

Figure 2 displays the contingency table graphically and states the adjusted contingency coefficient, the p-value, and the effect size. The percentage of well-structured tasks is statistically significant, $p = .011$, and large, $w = .89$. Thus, we conclude that groups are likely to exhibit CI when they work on sets of well-structured tasks but are unlikely to exhibit CI when they work on sets of ill-structured tasks. Task structure influences the presence or absence of a CI factor.

Due to the 10 different percentages of well-structured tasks treated as discrete categories in this contingency analysis, data in the contingency table is sparse. As a robustness check, we grouped somewhat similar percentages and

re-ran the analysis. First, we rounded percentages to the nearest multiple of 10% (i.e., we grouped 67% with 71%, and we grouped 75% with 83%), resulting in eight categories. The effect was significant, $p = .003$ and large, $w = .89$. Second, we grouped percentages as “more than 50%” and “at most 50%,” resulting in a two-by-two contingency table. Again, the effect was significant, $p = .000$ and large, $w = .82$. Thus, the finding that structure influences the presence or absence of a CI factor remains intact rather than being an artifact of the scale level. We performed an additional robustness check to see the extent to which the miscoding of tasks could create errors. See Online Appendix B for details.

4.3 | Analyses of other factors

CI studies have commonly studied groups of different sizes (midpoint ranging from 3 to 5; treated as a continuous variable), working in different cultural contexts (North America, Europe, Asia, Australia; a categorical variable), with different levels of collocation (same place, locally dispersed; treated as a binary dummy variable with 1 = locally dispersed), using different collaboration media (Face-to-Face, Chat, Various media, and Video and audio; a categorical variable), under different payment schemes (equal payment, performance-based payment; treated as a binary dummy variable with 1 = performance-based payment), and in different group contexts (ad-hoc randomly assigned groups for the CI experiment, intact teams collaborating beyond the CI experiment; treated as a binary dummy variable with 1 = intact). These variables may also affect whether groups exhibit CI (see Table 2 for the characterization of the studies). We examined the individual effects of these six variables and conducted contingency analyses for each of them. Results are presented in Online Appendices C through H and show no significant relationships for any of these six variables.

We also included these same six variables and the percentage of well-structured tasks (treated as a continuous variable) in a logistic regression to predict if groups do or do not exhibit CI (treated as a dependent binary dummy variable with 1 = yes). This analysis failed to produce a result, terminating with a quasi-complete separation error. A quasi-complete separation indicates that one of the predictor variables near-perfectly separates the outcome variable into its two outcomes, which results in an infinite coefficient (Albert & Anderson, 1984).

We, therefore, conducted a multiple regression analysis using SPSS GLM, which can be used to estimate binary outcomes (see Table 4). The resulting model found only the percentage of well-structured tasks to be statistically significant, $F(1,9) = 9.49$, $p = .013$; no other variable was significant. The effect size for the task variable was very large, partial $\eta^2 = .51$; J. Cohen (1988) considers a partial η^2 of .14 to be “large.” Thus, we conclude that the task variable was the cause of the separation error in the logistic regression. The model has a large effect size, with an R^2 of 71.7%. Variance inflation factors for all variables were less than 2, indicating there was no multicollinearity. As a robustness check, we included a weighting factor to account for the different number of participants in each study, and the statistical conclusions were the same.

5 | RQ2: ANALYSIS AND RESULTS

Because the initial results showed differences between well-structured and ill-structured tasks, we moved to RQ2 and conducted a meta-analysis to investigate how groups exhibit CI between differently structured tasks. The meta-analysis includes two central steps. The first step was the identification of primary studies. Second step was the collection of correlation coefficients between tasks in those studies, which are pooled into a combined correlation matrix which will serve as the new database for further analyses. Each of these steps is outlined in detail below.

TABLE 4 Regression results for the exhibition of CI

Effect	Std.Beta	Beta	Std. Error	95% Confidence Interval		p-value
				Lower Limit	Upper Limit	
Intercept	.00	-.75	1.31	-3.70	2.21	.582
Percentage of well-structured tasks	.67	1.24	.39	.35	2.14	.012
Group size (midpoint of range)	.18	.15	.44	-.85	1.15	.744
Cultural context (base case is Europe)						
North America	.35	.34	.41	-.58	1.27	.419
Australia	-.12	-.28	.54	-1.49	.94	.619
Asia	.04	.07	.42	-.89	1.03	.873
Collocation	-.09	-.15	.41	-1.07	.78	.730
Collaboration medium (base case is Face-to-Face)						
Video and audio	-.07	-.12	.42	-1.08	.84	.792
Chat	-.23	-.23	.31	-.92	.46	.471
Various media	-.08	-.13	.47	-1.20	.93	.781
Payment scheme	.23	.29	.40	-.62	1.19	.487
Group context	.22	.31	.64	-1.14	1.76	.640

Based on our original literature search for articles examining CI, we identified the studies that used quantitative analysis of CI and included correlation coefficients (r or ρ) between group task types. For studies missing correlation coefficients, we reached out to the authors. In total, our meta-analytic sample included 907 groups (N) from 17 studies (k) marked in the last column of Table 2. We used the meta-analytical approach from Wilson et al. (2016) to synthesize correlation coefficients to a single pooled correlation matrix. This approach extends traditional meta-analytical techniques as it deals with “complex datasets,” that is, datasets with more than one measure for a primary variable within studies and multiple measures of a primary variable across studies—both relevant issues in our data set.

Each primary study includes a correlation matrix stating the correlation coefficients of tasks. Hence, the correlation matrices of the studies were synthesized to a pooled 8×8 correlation matrix referencing McGrath’s (1984) eight task types. For example, Barlow and Dennis (2016b, GCS treatment, Table D2 in Online Appendix) and Engel et al. (2015, German study, Table 2) report raw correlation coefficients of .254 and .24 between a creativity task (McGrath’s, 1984; task type number 2) and a decision making task (McGrath’s, 1984; task type number 4). Both raw correlation coefficients address the same cell in the pooled correlation matrix. Details on correlation coefficients of primary studies are provided in Online Appendix I. In order to synthesize the raw correlation coefficients (and further raw correlation coefficients in this cell), we applied a “three-level multivariate mixed-effects weighted meta-regression model.” This model accounts for statistical dependencies associated with clustering units within levels, including those resulting from multiple tasks within the same studies (Wilson et al., 2016). As the following equation illustrates, the dependent variable r_{ik} represents the observed correlation coefficients $i = 1 - 267$ from the 17 studies (k). Each cell of the future pooled correlation matrix is represented with a unique independent dummy variable ($Cell_{1ik}, \dots, Cell_{28ik}$), which takes a value of 1 if coefficient i from study k is assigned to that cell and a value of 0 otherwise. This assigns each effect size stated in the primary correlation matrices to its “right position” in the pooled matrix.

$$r_{ik} = \beta_1 Cell_{1ik} + \beta_2 Cell_{2ik} + \dots + \beta_{28} Cell_{28ik} + \nu_{0k} + \eta_{ik} + \epsilon_{ik}.$$

TABLE 5 Pooled correlation matrix

Well or Ill	McGrath task type	Ill		Tendentially well	Tendentially ill	Ill		Well	Well
		1	2	3	4	5	6	7	8
Ill	1 Planning task	1							
Ill	2 Creativity task	.06	1						
Tendentially well	3 Intellectual task	.24	.36	1					
Tendentially ill	4 Decision-making task	.12	.14	.10	1				
Ill	5 Cognitive conflict task	.07	.17	.15	-.01	1			
Ill	6 Mixed-motive task	.12	.22	.21	.12	.12	1		
Well	7 Contests/battles	.24	.28	.20	.07	.10	.11	1	
Well	8 Performances	.36	.30	.25	.12	.15	.16	.26	1

Note. $k = 17$.

The use of a no-intercept model permits interpreting the respective regression coefficients as pooled correlation coefficients. 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 cells in the matrix, Level 3 random effects capture random effects of all correlation coefficients in the matrix. The estimation error ε_{ik} is assumed to be normally distributed with a mean of 0 and variance of ν_{ik} . The unconditional Level 2 random effects approximate the conditional sampling covariance between observed correlations from the same study. We assume that errors at different levels are uncorrelated. As suggested by Wilson et al. (2016), we used the statistical software R package *metafor* to conduct the analysis. The pooled correlation matrix resulting from the meta-analysis is displayed in Table 5.

The average correlation between pairs of well-structured tasks was .24. In contrast, the average correlation between pairs of ill-structured tasks was .11, indicating that performance on an ill-structured task was only slightly correlated with performance on other ill-structured tasks. We see high correlations among the creativity tasks (i.e., brainstorming) and the well-structured tasks. As discussed above, although brainstorming has ill-structured traits, it is unique in that it requires very little coordination.

5.1 | Exploratory factor analysis

Based on the pooled correlation matrix resulting from meta-analysis, we conducted an EFA to investigate the number of factors underlying the eight tasks, just as Woolley et al. (2010) did. We conducted principal axis factoring (PAF), often referred to as "common factor analysis" (Fabrigar et al., 1999). Parallel analysis was used to determine the number of factors to extract (Horn, 1965; Wood et al., 1996; Zwick & Velicer, 1982), and "promax" was used as it is the most common oblique factor rotation (Finch, 2006). Scholars interpret factor loadings $\geq .40$ as meaningfully related to a factor (Ford et al., 1986; Gefen & Straub, 2005).

We conducted two separate EFAs—one for well-structured tasks and one for ill-structured tasks. We used the statistical software R package *psych* and specifically the functions *fa* and *fa.parallel* to conduct the analysis. Tables 6 and 7 show the number of factors extracted and the associated proportions of variance explained for the two analyses. Task loadings above the threshold of $\geq .40$ appear in bold. For well-structured tasks, the parallel analysis yields a single general CI factor accounting for variance in the groups' performance scores on different well-structured tasks. Conversely, the cumulative empirical evidence does not support the exhibition of a single general CI factor for ill-structured tasks.

TABLE 6 EFA loading matrix for well-structured tasks

Well or Ill	McGrath task type	Factor 1
Tendentially well	3 Intellective task	.44
Well	7 Contests/battles	.45
Well	8 Performances	.57
Proportion of variance explained		.24

Note. $N = 907$. Factor loadings $< .40$ are in boldface.

TABLE 7 EFA loading matrix for Ill-structured tasks

Well or Ill	McGrath task type	Factor 1	Factor 2	Factor 3	Factor 4
Ill	1 Planning task	.01	.06	.08	.37
Ill	2 Creativity task	.35	.18	.18	-.16
Tendentially ill	4 Decision-making task	-.01	-.07	.46	.10
Ill	5 Cognitive conflict task	.00	.46	-.07	.08
Ill	6 Mixed-motive task	.52	-.06	-.07	.10
Proportion of variance explained		.08	.06	.06	.04

Note. $N = 907$. Factor loadings $< .40$ are in boldface.

6 | DISCUSSION

The current state of research on CI presents mixed results in that some studies have found groups to exhibit CI while others have not. Our primary goal was to understand what influences whether groups exhibit CI. Because CI is a measure of a group's ability to perform consistently well across tasks, we examined the fundamental structure of tasks used in prior research to understand disparate findings. Our analyses show that the structure of the tasks has both a significant and large effect on whether groups exhibit CI. For well-structured tasks, CI has a single general factor structure. Conversely, when groups are presented with a set of mostly ill-structured tasks, they do not exhibit CI in the manner defined by Woolley et al. (2010). Hence, their performance is not consistent across tasks. Given that task structure matters, our secondary goal was to understand the pattern of factors groups exhibit for well-structured or ill-structured tasks. Our findings show a single general CI factor for well-structured tasks but not for ill-structured tasks. Our results suggest that ill-structured tasks have a multi-factor pattern. Each of the multiple factors reflects particular tasks of McGrath's (1984) typology, but the sparse data strongly influence this result. With more data, we may see a different pattern.

We believe this study has three important theoretical implications. First, our results offer a theory-based reason for the currently mixed results in CI literature where some groups exhibit CI, and some groups do not. We conclude that there has been a mismatch between the theory and measurement of CI that has led to inconsistencies in prior research results. The original definition of the CI construct was too all-encompassing, referring to a group's ability to perform consistently well across a variety of (well-structured and ill-structured) tasks (Woolley et al., 2010). This original definition left room for researchers to use different tasks in their experiments, which yielded inconsistent findings. This study shows an underlying factor—task structure—that has led to these inconsistencies. Hence, the inconsistencies might have emerged not purely from a theoretical or a measurement issue but how research has matched the theoretical definition to its measurement.

Second, our results suggest that we should narrow the current definition of CI to a group's ability to perform consistently well across a variety of well-structured tasks. With that definition, we can continue to measure CI as done by Woolley et al. (2010). The original study by Woolley et al. (2010) used a set of primarily well-structured tasks to

measure CI. This approach was adopted by most researchers, who found similar results for a general CI factor. Existing theories and research of CI still apply but are limited in scope to well-structured tasks. One important next step for CI research is thus to use this aligned version of the theoretical definition of CI and its measurement so that different researchers share a common understanding.

Third, one important question is whether we should conclude that groups do not exhibit CI across ill-structured tasks or whether we should reconceptualize CI for ill-structured tasks as being a multi-dimensional phenomenon, similar to research on individual intelligence. The definition and measurement of general individual intelligence have been among the most researched topics in psychology over the last 100 years (Deary, 2000). Currently, individual intelligence is theorized and measured on three different levels, namely psychometric, psychological, and social (Davidson & Kemp, 2011; Eysenck, 1988; Flynn, 2007). As Woolley et al. (2010) defined and measured CI in analogy with Spearman's "psychometric *g*," the current CI discourse centers on the psychometric level. For individual intelligence, Spearman (1904) was first to suggest one general factor (along with some less important test-specific factors), while Thurstone and Thurstone (1941) were first to suggest seven broad factors which could be interpreted as forms of intelligence (e.g., verbal comprehension and number facility).

Current psychometric models of individual intelligence somewhat reconcile this discrepancy via hierarchical models of individual intelligence, suggesting that the existence of *g* does not negate the simultaneous coexistence of other specific forms of intelligence (e.g., Gf-Gc theory, three-stratum theory, or CHC theory of cognitive abilities; Davidson & Kemp, 2011). In analogy with this development, groups may simultaneously exhibit a general form of CI (i.e., a *c* factor) and specific forms of CI across a wide variety of well-structured and ill-structured tasks. For example, we may need to think of negotiation CI, planning CI, creative CI, or other forms of CI reflecting tasks of McGrath's (1984) circumplex (see first three columns in Table 1) or other group task typologies (e.g., S. G. Cohen & Bailey, 1997; Driskell et al., 1987; Laughlin & Ellis, 1986; McGrath, 1984; Schraw et al., 1995; Steiner, 1972; Straus, 1999; Wildman et al., 2012).

6.1 | Limitations

Our study has several limitations. First, our sample size is small, with 21 studies reporting data from about 5300 individuals in 1419 groups (with 17 studies reporting data from 3218 individuals in 907 groups used in the meta-analysis). While more data would be better, this sample resulted from a comprehensive literature review consisting of three different search strategies. It represents the cumulative empirical evidence to date. Like Credé et al. (2017), we are limited by the data currently available. The sample size primarily affects the EFA of the pooled correlations of ill-structured tasks, where the empirical evidence available is too sparse to draw definite conclusions. We encourage researchers to redo these analyses once more studies have been published.

Associated with the small sample size, publication bias against studies that find no significant results is a common concern in meta-analyses (also known as the file-drawer problem). To deal with potential publication bias, we created a funnel plot depicted in Online Appendix J. The plot revealed some asymmetry near the bottom of the plot, which potentially points to a sort of publication bias. However, it is also possible that the asymmetry in the plot roots in true heterogeneity between studies, as the average effect size per primary study (i.e., the average correlation between tasks in a primary study) depends on whether the tasks were ill- or well-structured. We further note that studies finding no significant evidence of CI have been published (8 of 21 studies [38%] of our sample found such null results), but we cannot eliminate the concern that the published evidence might be a biased subset of the full evidence. We reached out to authors of all published studies on CI and asked for unpublished studies by themselves or other authors. This query did not reveal further studies beyond the ones included in our analysis. As an additional measure, we calculated Orwin's (1983) fail-safe *N* as 17. Thus, there would need to be 17 undiscovered studies with an average effect size of zero that could be added to our analysis before the true effect size would be reduced to the smallest meaningful effect size. This figure is large compared to our current meta-analysis sample of also 17 studies.

Nevertheless, more data would be better, and we encourage researchers to continue with primary empirical studies on CI.

Our EFA of ill-structured tasks is limited by the 17 studies we could locate. Our results show a four-factor structure, but we do not conclude that this is indeed the exact number of factors or their content. Instead, we conclude that the cumulative empirical evidence available to date refutes the prevailing theory of a single CI factor for ill-structured tasks and suggests a more nuanced multi-factor structure. There may be even more specialized forms of CI in ill-structured task contexts in comparable ways as some authors conceptualize specialized forms of individual intelligence. We need future research to explore CI in ill-structured tasks. Researchers should aim at further understanding this factor structure, its antecedents, and ways to foster CI.

We examined whether other variables affected whether groups exhibit CI. The variables we included are limited in scope and granularity to factors reported in prior research. For example, information on variables thought to impact CI, such as task completion times, groups' social sensitivity, or percentage of females in the groups (Hyatt & Ruddy, 1997; Lemoine & Blum, 2019; Morgeson et al., 2005) is reported in some studies but missing in others. Future research should investigate a broader range of variables that were not included in the studies available to us but could potentially affect CI, such as team tenure (Hackman, 1992), team hierarchy (Greer et al., 2018), and personality factors, values, and attitudes (Bell, 2007).

Woolley et al. (2010) and all subsequent studies we analyzed here consider CI in analogy to the *g* factor of individual intelligence on a psychometric level. Hence, analyzing this accumulated evidence, our study is also limited to residing on the psychometric level. While individual intelligence research started at the psychometric level, it did not end there. As noted before, intelligence is not the same as *g*. Continuing the analogy with individual intelligence and in the notion of our findings pointing to a multi-factor structure of CI, this discourse may move forward and analyze analogies between individual intelligence and CI on the psychological and social levels. Theories at the psychological level aim at determining the neural basis of individual intelligence (Davidson & Kemp, 2011). Theories at the social level "view intelligence as a complex dynamic system involving interactions between mental processes, contextual influences, and multiple abilities" (Davidson & Kemp, 2011, p. 67). An example is the triarchic theory of successful intelligence proposing three types of individual intelligence (creative, analytic, and practical intelligence; Sternberg, 1985). To deepen our understanding of CI, future research should aim at theorizing CI at these levels and analyze its basis and social embeddedness.

6.2 | Implications for future research

This work provides several implications for future research. Our findings show that there is something about how groups work together in well-structured tasks that enables them to exhibit CI. Engaging in certain group activities that fit well-structured tasks enables groups to perform consistently well across tasks. We believe this constitutes a good starting point for future research as we should investigate how groups work on well-structured tasks to understand what enables them to exhibit CI.

We also need systematic research to understand how groups approach ill-structured vs. well-structured tasks. For example, groups show behaviors (Klein et al., 2004), and researchers may conduct interaction process analysis of literal content and process meaning (Bales, 1950; Kelly, 2000) to develop process models revealing the core activities and group interaction behaviors over time (Ancona et al., 2001). Activities and interaction behaviors underlying group performance on ill-structured tasks may then be compared to the activities and interaction behaviors that drive group performance on well-structured tasks. The core activities and interaction behaviors that influence performance on well-structured tasks may not be those that influence performance on ill-structured tasks. In ill-structured tasks, groups might apply different core activities or different interaction behaviors. This thought is to some degree analogous to research on individual intelligence. Spearman and other scholars saw that a psychological or even neurobiological explanation of *g* was needed (Mackintosh, 2011). This led to Sternberg (1977) and Pellegrino (1986) setting out to

TABLE 8 Questions in the field of CI for future research

Category	Questions
CI - Theory and Measurement	<ul style="list-style-type: none"> • If we narrow the theoretical definition of CI to well-structured tasks, is CI really CI, or does doing so make it something else? • How can the theoretical definition of CI be aligned with its measurement so that different researchers share a common understanding? • How many facets (i.e., factors) does CI have across ill-structured tasks? • Which facets of CI did prior research applying mostly ill-structured tasks find?
CI across ill-structured tasks—analogy with individual intelligence	<ul style="list-style-type: none"> • How do facets of CI across ill-structured tasks align with established research in individual intelligence? • Which models of individual intelligence (e.g., Gf-Gc theory, three-stratum theory, or CHC theory of cognitive abilities) may best be applicable in CI in ill-structured task contexts? • How can CI be theorized beyond psychometric levels, that is, on psychological and social levels?
Variables correlated with CI	<ul style="list-style-type: none"> • How is CI in different contexts affected by further study level characteristics, such as completion times, groups' social sensitivity, or percentage of females in the group? • How do study level characteristics shown to be correlated with CI in well-structured tasks relate to CI in ill-structured tasks? • Can we use goal setting or leadership behaviors to make ill-structured tasks even modestly more structured? If so, do CIs that are potentially unleashed differ in nature? • Given that prior research finds CI in well-structured tasks to be predictive of performance on more complex tasks (see Woolley et al., 2010), what outcomes could be predicted by various facets of CI on ill-structured tasks?
Group processes underlying CI	<ul style="list-style-type: none"> • What core activities and interaction behaviors underly groups exhibiting or not exhibiting CI across well-structured or ill-structured tasks, and do these core activities differ? • What might interventions or training look like to help groups overcome the difficulties associated with ill-structured tasks to exhibit CI? • Given that no single factor emerged in ill-structured tasks, are there certain processes or variables that predict performance on some types of ill-structured tasks but not others?

understand the “cognitive components” of reasoning underlying individual intelligence and Anderson (1992) starting a search of “elementary cognitive tasks” underlying individual intelligence.

Such research on CI would likely require theoretical work to discern the core group activities at a level below available group task taxonomies. For groups, this might be easier than for individuals, as group processes are arguably easier to observe and measure than individual mental processes. Drawing a wide analogy, what we are suggesting is a kind of process mining (van der Aalst, 2012a, 2012b). Process mining—a major trend in computer science, information systems, and business process management—is a set of analytic data techniques that uses event logs to identify processes. Thus, we need additional empirical studies (likely lab experiments) with the in-depth recording of events.

Research could also devise interventions to help groups overcome the difficulties associated with ill-structured tasks. One approach may be providing tools that make ill-structured tasks look more like well-structured tasks by providing an analysis plan, as Barlow and Dennis (2016a) did. Another intervention could be training to help groups engage in certain activities in a specific sequence that helps them overcome boundaries to exhibit CI.

Table 8 summarizes the most important questions we see for future research that should be explored in order to advance our current understanding of CI.

6.3 | Implications for practice

CI has a profound impact on the creation and functioning of groups in organizations. The findings of our study suggest that groups often exhibit CI (i.e., consistent performance) across well-structured tasks. Groups which do well on one well-structured task are likely to do well on other well-structured tasks because the work processes or activities used for one well-structured task are likely to transfer to other well-structured tasks. Thus, using past performance to predict future performance on well-structured tasks is a viable strategy for organizations when assigning groups to tasks. In well-structured task settings, practical insights from those studies finding a CI factor still apply. In our sample, factors not showing significance as boundary conditions in our analysis were group size (up to about five members), different cultural contexts, collocation, collaboration medium, payment scheme, and group context. For example, groups working face-to-face are as likely to exhibit CI as are groups working remotely over audio and video (e.g., Zoom), as are groups in Europe or the U.S. However, the data did not enable us to test the effects of changes to these factors (e.g., groups that establish CI working face-to-face, which then transition to using Zoom).

Further, groups do not exhibit CI across sets of tasks that are ill-structured. Groups that do well on one ill-structured task are not necessarily more likely to do well on a different ill-structured task. We caution practitioners that relying on groups' past performance when assigning groups to ill-structured tasks will not necessarily lead to the best performance. Specifically, findings from our meta-analysis show that the cumulative empirical evidence available to date even suggests a more nuanced multi-factor structure in ill-structured task contexts. Although empirical evidence available is too sparse to draw definite conclusions, when assembling a group for a set of ill-structured tasks (e.g., often innovative, and creative tasks which might occur in an early concept stage in organizational projects), practitioners may carefully consider whether it is possible to create a structure to help groups overcome the difficulties associated with ill-structured tasks.

7 | CONCLUSION

Our research shows the percentage of well-structured tasks enables or inhibits CI. Groups exhibit CI when they work on well-structured tasks but not on ill-structured tasks. Groups that do well on some well-structured tasks are likely to do similarly well on other well-structured tasks. Our main contributions are threefold. First, we clarify the boundary conditions for CI, including factors that are boundary conditions (like task structure) and those that are potentially not (e.g., group size). Second, in finding the boundary condition of task structure, we offer new theoretical insight and propose refining the CI definition and measurement. Finally, we articulate future research to move CI literature forward, particularly with regard to ill-structured tasks.

ACKNOWLEDGMENTS

This research was supported in part by the German Research Foundation (DFG; Grant Number 343128888). Open Access funding enabled and organized by Projekt DEAL.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article.

ORCID

Valerie Graf-Drasch  <https://orcid.org/0000-0002-2338-8373>

REFERENCES

Aggarwal, I., Woolley, A. W., Chabris, C. F., & Malone, T. W. (2019). The impact of cognitive style diversity on implicit learning in teams. *Frontiers in Psychology*, 10, Article 112. <https://doi.org/10.3389/fpsyg.2019.00112>

- Agresti, A., & Kateri, M. (2011). *Categorical data analysis*. Wiley.
- Albert, A., & Anderson, J. A. (1984). On the existence of maximum likelihood estimates in logistic regression models. *Biometrika*, 71(1), 1–10. <https://doi.org/10.1093/biomet/71.1.1>
- Alper, S., Tjosvold, D., & Law, K. S. (2000). Conflict management, efficacy, and performance in organizational teams. *Personnel Psychology*, 53(3), 625–642. <https://doi.org/10.1111/j.1744-6570.2000.tb00216.x>
- Ancona, D. G., Okhuysen, G. A., & Perlow, L. A. (2001). Taking time to integrate temporal research. *Academy of Management Review*, 26(4), 512–529. <https://doi.org/10.5465/amr.2001.5393887>
- Anderson, M. (1992). *Intelligence and development: A cognitive theory*. Oxford University Press.
- Bales, R. F. (1950). *Interaction process analysis: A method for the study of small groups*. Addison-Wesley Press.
- Barlow, J. B. (2017). Collective intelligence and its relationship to collective individual intelligence. In *Collective Intelligence Conference*.
- Barlow, J. B., & Dennis, A. R. (2016a). Group coordination structuring: A process to facilitate collective intelligence in virtual group work. In *Collective Intelligence Conference*.
- Barlow, J. B., & Dennis, A. R. (2016b). Not as smart as we think: A study of collective intelligence in virtual groups. *Journal of Management Information Systems*, 33(3), 684–712. <https://doi.org/10.1080/07421222.2016.1243944>
- Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology*, 83(3), 377–391. <https://doi.org/10.1037/0021-9010.83.3.377>
- Bates, T. C., & Gupta, S. (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. <https://doi.org/10.1016/j.intell.2016.11.004>
- Bell, S. T. (2007). Deep-level composition variables as predictors of team performance: A meta-analysis. *The Journal of Applied Psychology*, 92(3), 595–615. <https://doi.org/10.1037/0021-9010.92.3.595>
- Campion, M. A., Medsker, G. J., & Higgs, A. C. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. *Personnel Psychology*, 46(4), 823–847. <https://doi.org/10.1111/j.1744-6570.1993.tb01571.x>
- Chikersal, P., Tomprou, M., Kim, Y. J., Woolley, A. W., & Dabbish, L. (2017). Deep structures of collaboration: Physiological correlates of collective intelligence and group satisfaction. In *Twentieth ACM Conference on Computer Supported Cooperative Work and Social Computing*.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Cohen, S. G., & Bailey, D. E. (1997). What makes teams work: Group effectiveness research from the shop floor to the executive suite. *Journal of Management*, 23(3), 239–290. <https://doi.org/10.1177/014920639702300303>
- Credé, M., & Howardson, G. N., (2017). The structure of group task performance—A second look at “collective intelligence”: Comment on Woolley et al. (2010). *Journal of Applied Psychology*, 102(10), 1483–1492. <https://doi.org/10.1037/apl0000176>
- Davidson, J. E., & Kemp, I. A. (2011). Contemporary models of intelligence. In (R. J. Sternberg & S. B. Kaufman Eds.), *The Cambridge Handbook of Intelligence* (pp. 58–82). Cambridge University Press.
- Dearry, I. (2000). *Looking down on human intelligence: From psychometrics to the brain*. Oxford University Press.
- Demetriou, A. (2002). Tracing psychology's invisible giant and its visible guards. In (R. J. Sternberg & E. Grigorenko Eds.), *The general factor of intelligence: How general is it?* (pp. 3–18). Psychology Press.
- Diehl, M., & Stroebe, W. (1987). Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 53(3), 497–509. <https://doi.org/10.1037/0022-3514.53.3.497>
- Driskell, J. E., Salas, E., & Hogan, R. (1987). A taxonomy for composing effective naval teams (Technical Report Number 87-002). Orlando, FL: US Naval Training Systems Center Technical Reports.
- Engel, D., Woolley, A. W., Aggarwal, I., Chabris, C. F., Takahashi, M., Nemoto, K., Kaiser, C., Kim, Y. J., & Malone, T. W. (2015). Collective intelligence in computer-mediated collaboration emerges in different contexts and cultures. In *Thirty-third Conference for Human-Computer Interaction (HCI)*.
- Engel, D., Woolley, A. W., Jing, L. X., Chabris, C. F., & Malone, T. W. (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), 1–16. <https://doi.org/10.1371/journal.pone.0115212>
- Eysenck, H. J. (1988). The concept of “intelligence”: Useful or useless? *Intelligence*, 12, 1–16. [https://doi.org/10.1016/0160-2896\(88\)90019-0](https://doi.org/10.1016/0160-2896(88)90019-0)
- Fabrigar, L. R., Wegener, D. T., & MacCallum, R. C. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Finch, H. (2006). Comparison of the performance of varimax and promax rotations: Factor structure recovery for dichotomous items. *Journal of Educational Measurement*, 43(1), 39–52. <https://doi.org/10.1111/j.1745-3984.2006.00003.x>
- Fischhoff, B., & Broomell, S. B. (2020). Judgment and decision making. *Annual Review of Psychology*, 71, 331–355. <https://doi.org/10.1002/wcs.65>

- Flynn, J. R. (2007). *What is intelligence? Beyond that Flynn effect*. Cambridge University Press.
- Ford, J. K., MacCallum, R. C., & Tait, M. (1986). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Personnel Psychology*, 39(2), 291–314. <https://doi.org/10.1111/j.1744-6570.1986.tb00583.x>
- Galton, F. (1907). Vox populi. *Nature*, 75, 450–451.
- Gardner, H. E. (2008). *Multiple intelligences: New horizons in theory and practice*. Basic books.
- Ge, X., & Land, S. M. (2003). Scaffolding students' problem-solving processes in an ill-structured task using question prompts and peer interactions. *Educational Technology Research and Development* 51(1), 21–38. <https://doi.org/10.1007/BF02504515>
- Gefen, D., & Straub, D. (2005). The application of exploratory factor analysis in applied psychology: A critical review and analysis. *Communications of the Association for Information Systems*, 16(1), 91–109. <https://doi.org/10.1111/j.1744-6570.1986.tb00583.x>
- Gick, M. L. (1986). Problem-solving strategies. *Educational Psychologist*, 21(1-2), 99–120. <https://doi.org/10.1080/00461520.1986.9653026>
- Gimpel, H., & Graf-Drasch, V. (2020). Media synchronicity and collective intelligence. *Working Paper*.
- Greer, L. L., Jong, B. A. d., Schouten, M. E., & Dannals, J. E. (2018). Why and when hierarchy impacts team effectiveness: A meta-analytic integration. *The Journal of Applied Psychology*, 103(6), 591–613. <https://doi.org/10.1037/apl0000291>
- Gupta, P., Kim, Y. J., Gliksou, E., & Woolley, A. W. (2019). Digitally nudging team processes to enhance collective intelligence. In *Collective Intelligence Conference*.
- Hackman, J. R. (1992). Group influences on individuals in organizations. In (M. D. Dunnette & L. M. Hough Eds.), *Handbook of industrial and organizational psychology* (pp. 199–267). Consulting Psychologists Press.
- Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in Experimental Social Psychology*, 8, 45–99. [https://doi.org/10.1016/S0065-2601\(08\)60248-8](https://doi.org/10.1016/S0065-2601(08)60248-8)
- Hope, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(3), 582–598. <https://doi.org/10.1111/j.2517-6161.1968.tb00759.x>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Hu, J., & Liden, R. C. (2011). Antecedents of team potency and team effectiveness: An examination of goal and process clarity and servant leadership. *The Journal of Applied Psychology*, 96(4), 851–862. <https://doi.org/10.1037/a0022465>
- Hyatt, D. E., & Ruddy, T. M. (1997). An examination of the relationship between work group characteristics and performance: Once more into the breach. *Personnel Psychology*, 50(3), 553–585. <https://doi.org/10.1111/j.1744-6570.1997.tb00703.x>
- Jensen, A. R. (2002). Psychometric g: Definition and substantiation. In (R. J. Sternberg & E. Grigorenko Eds.), *The general factor of intelligence: How general is it?* (pp. 39–53). Psychology Press.
- Jonassen, D. H. (1997). Instructional design models for well-structured and ill-structured problem-solving learning outcomes. *Educational Technology Research and Development*, 45(1), 65–94. <https://doi.org/10.1007/BF02299613>
- Kelly, J. R. (2000). Interaction process analysis in task-performing groups. In (A. P. Beck & C. M. Lewis Eds.), *The process of group psychotherapy: Systems for analyzing change*. (pp. 49–65). American Psychological Association, <https://doi.org/10.1037/10378-003>
- Kim, Y. J., Engel, D., Woolley, A. W., Lin, J. Y.-T., McArthur, N., & Malone, T. W. (2017). What makes a strong team? Using collective intelligence to predict team performance in league of legends. In *Twentieth ACM Conference on Computer Supported Cooperative Work and Social Computing*.
- Kitchener, K. S. (1983). Cognition, metacognition, and epistemic cognition. *Human Development*, 26(4), 222–232. <https://doi.org/10.1159/000272885>
- Kitchener, K. S., & King, P. M. (1981). Reflective judgment: Concepts of justification and their relationship to age and education. *Journal of Applied Developmental Psychology*, 2(2), 89–116. [https://doi.org/10.1016/0193-3973\(81\)90032-0](https://doi.org/10.1016/0193-3973(81)90032-0)
- Kittur, A., Lee, B., & Kraut, R. E. (2009). Coordination in collective intelligence. In D. R. Olsen, R. B. Arthur, K. Hinckley, M. R. Morris, S. Hudson, & S. Greenberg (Chairs), *Twenty-seventh International Conference on Human Factors in Computing Systems*.
- Klein, K. J., Lim, B.-C., Saltz, J. L., & Mayer, D. M. (2004). How do they get there? An examination of the antecedents of centrality in team networks. *Academy of Management Journal*, 47(6), 952–963. <https://doi.org/10.5465/20159634>
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. <https://doi.org/10.2307/2529310>
- Larsen, K. R., Hovorka, D. S., Dennis, A. R., & West, J. D. (2019). Understanding the elephant: The discourse approach to boundary identification and corpus construction for theory review articles. *Journal of the Association for Information Systems*, 20(7), 887–927. <https://doi.org/10.17705/1jais.00556>
- Laughlin, P. R. (1986). Demonstrability and social combination processes on mathematical intellectual tasks. *Journal of Experimental Social Psychology*, 22(3), 177–189.

- Lemoine, G. J., & Blum, T. C. (2019). Servant leadership, leader gender, and team gender role: Testing a female advantage in a cascading model of performance. *Personnel Psychology*, 74(1), 3–28. <https://doi.org/10.1111/peps.12379>
- LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology*, 61(2), 273–307. <https://doi.org/10.1111/j.1744-6570.2008.00114.x>
- Mackintosh, N. J. (2011). History of theories and measurement of intelligence. In (R. J. Sternberg & S. B. Kaufman Eds.), *The Cambridge Handbook of Intelligence* (pp. 3–19). Cambridge University Press.
- Malone, T. W., & Bernstein, M. S. (2015). *Handbook of collective intelligence*. The MIT Press.
- Malone, T. W., & Crowston, K. (1990). What is coordination theory and how can it help design cooperative work systems? In (F. Halasz Ed.), *ACM Conference on Computer-Supported Cooperative Work (CSCW)* (pp. 357–370). ACM Press.
- Malone, T. W., & Crowston, K. (1994). The interdisciplinary study of coordination. *ACM Computing Surveys*, 26(1), 87–119. <https://doi.org/10.1145/174666.174668>
- McGrath, J. E. (1984). *Groups: Interaction and performance*. Prentice-Hall.
- Meslec, N., Aggarwal, I., & Curseu, P. L. (2016). The insensitive ruins it all: Compositional and compilational influences of social sensitivity on collective intelligence in groups. *Frontiers in Psychology*, 7, Article 676. <https://doi.org/10.3389/fpsyg.2016.00676>
- Morgeson, F. P., Reider, M. H., & Campion, M. A. (2005). Selecting individuals in team settings: The importance of social skills, personality characteristics, and teamwork knowledge. *Personnel Psychology*, 58(3), 583–611. <https://doi.org/10.1111/j.1744-6570.2005.655.x>
- Newell, A., & Simon, H. A. (1972). *Human problem solving*. Prentice-Hall.
- Ostrowski, B., Woolley, A. W., & Haan, A. K.-W. (2019). Group brainstorming: The effects of collective intelligence, individual ability, and task structure. In *Collective Intelligence Conference*.
- Ott, R. L., Larson, R., Rexroat, C., & Mendenhall, W. (1992). *Statistics: A tool for the social sciences* (5th ed.). PWS-Kent Publishing Company.
- Pellegrino, J. W. (1986). Deductive reasoning ability. In (R. J. Sternberg Ed.), *Human abilities: An information processing approach*. W.H. Freeman and Company.
- Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W., & Woolley, A. W. (2021). Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences of the United States of America*, 118(21). <https://doi.org/10.1073/pnas.2005737118>
- Rowe, L. I. (2019). *Exploring collective intelligence in human groups* [Unpublished Doctoral Thesis], Melbourne Graduate School of Education. <http://hdl.handle.net/11343/233593>
- Sandelands, L. E., & Stablein, R. E. (1987). The concept of organization mind. In (N. D. & S. B. Eds.), *Research in the sociology of organizations* (pp. 135–161). JAI Press.
- Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Cognitive processes in well-defined and ill-defined problem solving. *Applied Cognitive Psychology*, 9(6), 523–538. <https://doi.org/10.1002/acp.2350090605>
- Shea, G. P., & Guzzo, R. A. (1987). Groups as human resources. In (K. M. Rowland & G. R. Ferris Eds.), *Research in human resources and personnel management* (5th ed., pp. 323–356). JAI Press.
- Sheskin, D. (2011). *Handbook of parametric and nonparametric statistical procedures* (5th ed.). Chapman & Hall/CRC.
- Simon, H. A. (1973). The structure of ill structured problems. *Artificial Intelligence*, 4(3-4), 181–201. [https://doi.org/10.1016/0004-3702\(73\)90011-8](https://doi.org/10.1016/0004-3702(73)90011-8)
- Sinott, J. D. (1989). A model for solution of ill-structured problems: Implications for everyday and abstract problem solving. In (J. D. Sinott Ed.), *Everyday problem solving: Theory and applications* (pp. 77–99). Praeger Publishers.
- Spearman, C. (1904). General Intelligence" Objectively determined and measured. *The American Journal of Psychology*, 15(2), 201–292. <https://doi.org/10.1037/11491-006>
- Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Erlbaum.
- Steiner, I. D. (1972). *Group processes and productivity*. Academic Press: New York, NY.
- Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge University Press.
- Sulis, W. (1997). Fundamental concepts of collective intelligence. *Nonlinear Dynamics, Psychology, and Life Sciences*, 1(1), 35–53. <https://doi.org/10.1023/A:1022371810032>
- Sulis, W. (2009). Collective intelligence: Observations and models. In (S. J. Guastello, M. Koopmans, & D. Pincus Eds.), *Chaos and complexity in psychology: The theory of nonlinear dynamical systems* (pp. 41–72). Cambridge University Press.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations*. Doubleday.
- Straus, S. G. (1999). Testing a typology of tasks: An empirical validation of McGrath's (1984) group task circumplex. *Small Group Research*, 30(2), 166–187.

- Thurstone, L. L., & Thurstone, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs* (Vol 2). Chicago: University of Chicago.
- van der Aalst, W. (2012a). Process mining. *Communications of the ACM*, 55(8), 76–83. <https://doi.org/10.1145/2240236.2240257>
- van der Aalst, W. (2012b). Process Mining: Overview and opportunities. *ACM Transactions on Management Information Systems*, 3(2), 1–17. <https://doi.org/10.1145/2229156.2229157>
- Voss, J. F., & Post, T. A. (1988). On the solving of ill-structured problems. In (M. Chi, R. Glaser, & M. Farr Eds.), *The nature of expertise* (pp. 261–285). Lawrence Erlbaum Associates.
- Weick, K. E., & Roberts, K. H. (1993). Collective mind in organizations: Heedful interrelating on flight decks. *Administrative Science Quarterly*, 38(3), 357–381. <https://doi.org/10.2307/2393372>
- Wheeler, W. M. (1911). The ant-colony as an organism. *Journal of Morphology*, 22(2), 307–325. <https://doi.org/10.1002/jmor.1050220206>
- Wildman, J. L., Thayer, A. L., Rosen, M. A., Salas, E., Mathieu, J. E., & Rayne, S. R. (2012). Task types and team-level attributes: Synthesis of team classification literature. *Human Resource Development Review*, 11(1), 97–129. <https://doi.org/10.1177/1534484311417561>
- Williams, W. M., & Sternberg, R. J. (1988). Group intelligence: Why some groups are better than others. *Intelligence*, 12(4), 351–377. [https://doi.org/10.1016/0160-2896\(88\)90002-5](https://doi.org/10.1016/0160-2896(88)90002-5)
- Wilson, S. J., Polanin, J. R., & Lipsey, M. W. (2016). Fitting meta-analytic structural equation models with complex datasets. *Research Synthesis Methods*, 7(2), 121–139.
- Wood, J. M., Tataryn, D. J., & Gorsuch, R. L. (1996). Effects of under- and overextraction on principal axis factor analysis with varimax rotation. *Psychological Methods*, 1(4), 354–365. <https://doi.org/10.1037/1082-989X.1.4.354>
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688. <https://doi.org/10.1126/science.1193147>
- Woolley, A. W., Riedl, C., Kim, Y. J., & Malone, T. W. (2017). More evidence for a general collective intelligence factor in human groups: A meta-analysis. In *Collective Intelligence Conference*.
- Zwick, W. R., & Velicer, W. F. (1982). Factors influencing four rules for determining the number of components to retain. *Multivariate Behavioral Research*, 17(2), 253–269. https://doi.org/10.1207/s15327906mbr1702_5

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Graf-Drasch, V., Gimpel, H., Barlow, J. B., & Dennis, A. R. (2022). Task structure as a boundary condition for collective intelligence. *Personnel Psychology*, 75, 739–761. Advance online publication. <https://doi.org/10.1111/peps.12489>