

Collective Intelligence: Concept, Mechanism, and Measurement

Authors: Chu Gaohong, Wang Zhimou, Hu Jing, Zhan Peida, Peida Zhan

Date: 2025-12-02T14:14:01+00:00

Abstract

Collective intelligence is the team-level general cognitive ability whereby teams collaboratively address complex tasks or solve problems through collaboration, communication, and knowledge sharing; its essence lies in transcending individual limitations to achieve group-level cognitive synergy and performance enhancement. However, current research in this field still faces dual challenges of conceptualization and measurement: conceptually, the coexistence of multidisciplinary perspectives leads to inconsistent conceptual definitions, and theories such as shared mental models, transactive memory systems, and interactive team cognition lack an integrative framework; methodologically, the evaluative paradigm excels at measuring overall effectiveness but is weak in revealing mechanisms, while the diagnostic paradigm is strong in process analysis but lacks ecological validity. This study systematically reviews the conceptual evolution of collective intelligence, elucidates the major theoretical models of its formation mechanisms, and compares evaluative and diagnostic measurement paradigms. On this basis, it proposes that future research should promote the integration of measurement paradigms, construct multimodal dynamic assessment systems, and strengthen research on human-AI collaborative teams, thereby expanding the theoretical boundaries and application prospects of collective intelligence.

Full Text

Preamble

Collective Intelligence: Conceptualization, Mechanism, and Measurement

Chu Gaohong¹, Wang Zhimou^{1,2}, Hu Jing¹, Zhan Peida¹

(1 Educational Neuroscience and Intelligent Measurement Laboratory, School of Psychology, Zhejiang Normal University, Jinhua 321004, China)

(2 Faculty of Psychology, Beijing Normal University, Beijing 100875, China)

Abstract: Collective intelligence refers to a team-level general cognitive ability through which team members collaborate, communicate, and share knowledge to tackle complex tasks or solve problems. Its essence lies in transcending individual limitations to achieve cognitive synergy and efficacy enhancement at the group level. Despite growing scholarly interest, the field continues to grapple with significant challenges in both conceptual clarity and methodological rigor. Conceptually, the construct is characterized by definitional heterogeneity stemming from diverse disciplinary perspectives; moreover, prominent theoretical frameworks—such as Shared Mental Models, Transactive Memory Systems, and Interactive Team Cognition—remain fragmented, lacking a unifying integrative model. In terms of measurement approaches, existing assessment approaches exhibit complementary strengths and limitations: while evaluative methods effectively capture overall team performance, they offer limited insight into underlying cognitive processes; conversely, diagnostic approaches provide detailed process-oriented analyses but often sacrifice ecological validity. This study systematically reviews the conceptual evolution of collective intelligence, explains the main theoretical models of its formation mechanisms, and compares evaluative and diagnostic measurement paradigms. Based on this, we propose future directions including promoting the integration of measurement paradigms, constructing a multimodal dynamic assessment system, and strengthening research on human-AI collaborative teams to expand the theoretical boundaries and practical applications of collective intelligence.

Keywords: collective intelligence, team cognition, Human-AI collaboration

Classification Number: B848

Introduction

Throughout history, as humans have confronted uncertain environments and complex challenges, intelligence has remained the core capacity enabling adaptation, survival, and innovation. Since Binet and Simon pioneered intelligence measurement in 1905, intelligence has been regarded not only as a key predictor of academic and occupational achievement but also as a fundamental psychological resource driving social progress, technological development, and cultural evolution. Research on the measurement, structure, and function of intelligence constitutes one of the earliest and most enduring topics in modern psychology. However, many real-world complex problems and major breakthroughs cannot be solved by individuals alone; rather, they emerge from collaboration and the manifestation of collective wisdom among multiple individuals. Individual intelligence does not always automatically translate into effective group performance (Woolley et al., 2010), revealing the limitations of traditional intelligence theories in explaining multi-member system performance.

Against the backdrop of organizations and society increasingly relying on teams to solve complex tasks, researchers have begun to explore whether teams possess an intelligence analogous to individual intelligence—collective intelligence (also translatable as “team intelligence”). Recently, *Science* listed “How does group

intelligence emerge?” among the 125 most cutting-edge scientific questions worldwide, underscoring the scientific significance of collective intelligence research. In this vein, Woolley et al. (2010) provided the first empirical evidence for the objective existence of collective intelligence by adapting individual intelligence test paradigms, demonstrating that collective intelligence is a systematic synergistic effect that significantly predicts team performance beyond individual members’ intelligence. As a typical team-level construct, research has shown that collective intelligence’ s formation and development are based on both individual member characteristics (i.e., intelligence and personality) and, more importantly, on team cognitive structures and social coordination mechanisms dynamically generated through member interaction. This systemic team-level attribute manifests in processes such as information sharing, perspective integration, and conflict coordination (He et al., 2019; Leitão et al., 2022; Li et al., 2019; Woolley et al., 2010), with its core function deriving from team-level cognition shaped by member cognitive interactions rather than simple aggregation of individual traits.

Collective intelligence is a team-level construct that emerges through teamwork. Due to its complexity beyond individual intelligence, it faces challenges in conceptual definition, theoretical exploration of formation mechanisms, and methodological construction. The lack of unified conceptual definitions and formation mechanisms has led to significant divergence in measurement approaches, which not only weakens comparability across studies but also constrains theoretical and practical advancement in the field. Second, traditional teams consist of human members who rely on their respective expertise, cognitive abilities, and social skills to complete tasks together. However, with the rapid development of artificial intelligence technology, modern team composition is no longer limited to humans but increasingly evolves into human-AI hybrid teams. From smartphone voice assistants to enterprise-level intelligent decision-making systems, AI applications have become widespread, with AI roles transcending tool attributes to become organizational members with specific roles (e.g., coordinators, organizers) and functions (Gupta & Woolley, 2021; Siemon, 2022). Growing research and practice demonstrate that Human-AI Collaboration is not only technically feasible but also an inevitable trend in future organizational development (Wang et al., 2020). The State Council’ s “Opinions on Deeply Implementing the ‘AI Plus’ Action” explicitly proposes promoting transformation in philosophy and social science research methods toward human-machine collaboration and exploring new research organizational forms adapted to the AI era. This strategic orientation urgently requires us to deeply investigate the essential connotation, core components, and scientific measurement methods of collective intelligence in the context of Human-AI Collaboration, providing theoretical support and practical guidance for achieving systematic paradigm transformation. Meanwhile, domestic research on collective intelligence remains relatively scarce, which is not conducive to implementing the guiding requirements for student collaborative learning and team competency cultivation outlined in policy documents such as the “Opinions of the Ministry of Education and Eight

Other Departments on Accelerating Educational Digitalization” and “China Education Modernization 2035.” Against this backdrop, systematically studying collective intelligence, revealing its formation mechanisms, and establishing scientific measurement methods have become critical supports for promoting policy implementation and cultivating team collaboration competencies, while also providing theoretical foundations and practical pathways for constructing new cooperation models in “AI Plus” environments.

Accordingly, this study aims to systematically review the conceptual evolution of collective intelligence, deeply analyze three mechanism models of team collaboration, and examine the advantages and limitations of two mainstream measurement paradigms. Based on this analysis, we propose three research directions: (1) integrating measurement paradigms through innovative task design and multi-dimensional process data analysis to construct a comprehensive assessment framework that balances ecological validity and mechanism diagnosis; (2) establishing a multimodal dynamic assessment system that integrates behavioral, physiological, and neural data sources to achieve quantification of the entire team interaction process; and (3) systematically exploring research paradigms and methodologies for human-AI collaborative teams, focusing on frontier challenges such as cognitive division of labor, coordination mechanisms, and trust establishment between humans and intelligent agents.

2.1 Conceptual Evolution of Collective Intelligence

The development of the collective intelligence concept reflects a fundamental shift in human understanding of team cognitive capabilities. Its evolution spans centuries, moving from early philosophical speculation and sociological skepticism toward empirical research and interdisciplinary integration, becoming a cutting-edge topic at the intersection of management, computational science, social science, and ecology in contemporary times.

This exploration originated in the 18th-century Enlightenment with Condorcet’s “jury theorem,” which posited that if individuals could make independent judgments based on rationality and private information, group decisions made through aggregation mechanisms could outperform single individuals, laying a theoretical foundation for collective intelligence. However, 19th-century sociology generally held skeptical views. Scholars such as Gustave Le Bon, analyzing group behavior during the French Revolution, proposed the “crowd theory,” arguing that individuals in groups easily become irrational and emotional, with collective behavior often leading to the annihilation of reason and degradation of judgment (Yu et al., 2018). Until the late 19th century, Galton challenged this notion through the “ox-weight guessing” experiment, demonstrating that the median judgment of a group could surpass that of individual experts, while biological observations also revealed that ants, bees, and other organisms exhibit “superorganism intelligence” through collaboration, suggesting deep roots of collective intelligence in nature (Galton, 1907; Seely, 1995; Wheeler, 1922; Yu et al., 2018).

The latter half of the 20th century witnessed a significant shift from phenomenological description to conceptual definition and cross-disciplinary theoretical construction. Research gradually moved beyond early simple questioning of group rationality to deeply exploring the unique advantages and internal mechanisms of group decision-making. Mid-period research began systematically identifying the functional value of group wisdom, with Wechsler (1956) first attempting an abstract definition of collective intelligence, emphasizing its role as a systemic capability for groups to respond to external challenges and achieve common goals. By the end of the 20th century, Heylighen (1999) extended this theory to the management domain, validating the rationality of group decisions through case studies and statistical methods, while Conrardt and Roper (2003) provided mathematical support for the advantages of collective collaboration from an evolutionary perspective through mathematical modeling.

Entering the 21st century, collective intelligence research has entered a stage of refinement and diversification. Surowiecki (2004) brought the concept to public attention and proposed three core problem types that collective intelligence can solve—cognition, coordination, and cooperation—establishing a preliminary framework for the field. More critically, the research focus shifted from “whether teams possess intelligence” to “how teams generate or form intelligence.” Woolley et al. (2010) proposed that collective intelligence is a team’s ability to complete various tasks, emphasizing that its core lies in member interaction, knowledge sharing, and coordination processes. This perspective laid the foundation for understanding collective intelligence in psychology, highlighting the critical role of interactive processes in shaping team performance (Curşeu et al., 2015; Krafft, 2018; Woolley et al., 2010; Woolley et al., 2015). Kurvers et al. (2015) focused on how knowledge integration and collaboration mechanisms enhance problem-solving efficiency, defining collective intelligence as the ability of groups to outperform individuals on cognitive tasks. Palak and Wojtkiewicz (2021) proposed from a team innovation perspective that collective intelligence is an active intellectual cooperation process focusing on shaping and leveraging innovative capabilities, thereby distinguishing it from traditional teamwork. With information technology development, Narayanan et al. (2022) proposed a more modern definition, stating that collective intelligence is the process by which intelligent agents (including humans and software agents) distributed in network environments participate in decision-making through mutual communication and information sharing, combining their environmental information and feedback from other members to collaboratively complete tasks and achieve common goals, emphasizing the value of digital tools and collaboration platforms in enhancing team cognition.

Looking back at the entire developmental trajectory, collective intelligence research has evolved from philosophical and sociological speculation, through biological behavioral empiricism and mathematical modeling, to a technology-enabled, multidisciplinary era. The core consensus has become increasingly clear: collective intelligence transcends the mechanical sum of individual capabilities, and its essence lies in members forming a team-level cognitive capacity

through dynamic interactive processes such as collaboration, communication, and knowledge sharing to effectively respond to complex challenges and solve problems. Based on this, we propose that collective intelligence can be defined as: a team-level general cognitive ability through which team members collaborate, communicate, and share knowledge to jointly tackle complex tasks or solve problems. This definition captures three key characteristics of collective intelligence: (1) team-level cognitive capacity as its essential attribute; (2) realization through social interaction (communication and collaboration) among members; and (3) function-oriented toward solving complex tasks.

2.2 Mechanism Models of Collective Intelligence

Based on the IPO (Input-Process-Output) model (Hackman, 1987), collective intelligence formation involves three aspects: the input layer includes potential resources such as cognitive abilities, knowledge reserves, and personality traits possessed by individual team members; the process layer comprises dynamic mechanisms through which members integrate, transform, and apply individual resources via interactive processes such as communication, coordination, conflict management, and knowledge sharing; and the output layer represents the cognitive capabilities and performance outcomes that emerge from interaction processes, exceeding the simple sum of individuals and representing the team as an integrated system. However, merely defining collective intelligence or describing its external manifestations (i.e., quantity of tasks completed, speed of task completion) is insufficient to reveal its deep operational principles. To truly understand how teams achieve “the whole is greater than the sum of its parts,” effectively respond to complex tasks, and demonstrate exceptional team cognitive capabilities, we must deeply explore their internal cognitive mechanisms—that is, understanding how teams cooperate is key to understanding collective intelligence.

To better explain the cognitive mechanisms in team collaboration and task execution, researchers have proposed various mechanism models, among which the most well-known are Shared Mental Models (SMMs), which emphasize team members’ shared understanding of tasks and roles (Cannon-Bowers et al., 1993); Transactive Memory Systems (TMS), which focus on the distribution and retrieval mechanisms of expertise among members (Wegner et al., 1991; Woolley & Gupta, 2024); and the Interactive Team Cognition model (ITC), which emphasizes real-time interactive cognitive processes (Cooke et al., 2013). Each model explains how members achieve knowledge integration, coordinated action, and goal attainment through specific cognitive structures and interactive processes, revealing the “black box” of collective intelligence emergence. A comparative summary of these mechanism models is presented in Table 1 .

Shared Mental Models posit that team members’ shared knowledge structures enable them to form correct interpretations and expectations of team tasks, thereby coordinating their behaviors to meet team task requirements and other members’ needs (Cannon-Bowers et al., 1993; Klimoski & Mohammed, 1994; Bai

Xinwen & Wang Erping, 2004), focusing on alignment of key psychological constructs such as task understanding and situational interpretation. Researchers commonly use Pathfinder techniques, multidimensional scaling, cognitive mapping, and communication data analysis, employing metrics such as similarity, accuracy, and distribution to quantify shared mental models (He Guibing & Yang Qiong, 2006; Lin Xiaomin et al., 2014). For example, similarity between two members' mental models can be calculated using QAP correlation coefficients (Mathieu et al., 2000). Weller et al. (2014) supported that shared mental models enhance internal team communication and coordination, particularly the importance of rapid decision-making in high-pressure environments such as military or medical teams. Shared mental models effectively illuminate how teams respond to dynamic situations in high-pressure contexts, thereby enhancing problem-solving capabilities and overall performance.

Transactive Memory Systems focus on forming distributed cognitive networks through differential knowledge allocation, enabling team members to effectively assume responsibilities for knowledge storage and retrieval in specific domains (Wegner et al., 1991). This model emphasizes shared cognition among team members regarding “who knows what,” expanding the concept of team cognition by focusing on knowledge sharing and retrieval. Transactive memory systems are primarily evaluated through social network analysis, measuring structural quality by quantifying expertise distribution efficiency (i.e., centrality, connectivity metrics). Research indicates that a well-functioning transactive memory system leads to better performance because team members can rapidly access specialized knowledge (Mol et al., 2015; Zheng, 2012), particularly in complex collaborative scenarios such as medical and engineering teams, where clear cognition of expertise distribution among members reduces communication costs and enhances crisis response speed (i.e., immediate mobilization of member skills in operating room teams). The Transaction Systems Model-Collective Intelligence (TSM-CI) further extends this logic, proposing that collective intelligence requires dynamic adaptation through the synergistic interaction of Transactive Memory Systems (TMS), Transactive Attention Systems (TAS), and Transactive Reasoning Systems (TRS) (Woolley & Gupta, 2024). Among these, TMS focuses on knowledge complementarity and metamemory updating, TAS manages attention allocation and emergent collaboration, while TRS maintains member motivation through goal negotiation and priority alignment. The three systems collaborate according to environmental demands, with TMS and TAS dominating efficiency and TRS strengthening maintenance functions (Woolley & Gupta, 2024).

Interactive Team Cognition represents a breakthrough team cognition theory that transcends traditional individual cognitive aggregation perspectives. Its core lies in revealing how teams construct and update team cognition through dynamic social interaction, positing that team cognition is fundamentally about how team members interact and share information (Cooke et al., 2013; Cooke, 2015). This model emphasizes the fluidity of cognitive processes, demonstrating that cognition emerges from social interaction rather than being a superposition

of individual member traits. Even when individual members have relatively limited capabilities, teams can still achieve successful collaborative performance as long as they possess effective team interaction. Interactive Team Cognition shifts the focus from individual knowledge aggregation to collaborative processes that lead to effective team operations, suggesting that performance can be enhanced by designing team tasks that promote effective communication and timely information sharing, thereby creating more adaptive team environments (Cooke et al., 2013). Measurement based on Interactive Team Cognition must focus on interaction quality rather than knowledge stock, with researchers employing methods that concentrate on the quality of communication, coordination, and information exchange among team members and their collaborative dynamics (Cooke et al., 2013; Wildman et al., 2014). Specifically, commonly used metrics include: communication coordination, which assesses collaborative consistency among members by analyzing dialogue synchrony and complementarity (Gorman et al., 2020; Reitman et al., 2025); pattern stability, which captures dynamic regularities in team interaction by using hidden Markov models to identify transition probabilities between communication states (Demir et al., 2019; Reitman et al., 2025); response latency, which measures the timeliness of information feedback among members (i.e., standard deviation of average response time) to reflect real-time collaboration efficiency (Gorman et al., 2020); and redundancy, which calculates the proportion of repetitive information to reveal potential efficiency losses or resource waste in coordination processes (Wooldridge et al., 2019). These metrics collectively construct a measurement framework centered on interactive processes, enabling more precise capture of the dynamic characteristics of team cognition and providing empirical evidence for optimizing team collaboration.

The three mechanism models progress layer by layer from knowledge sharing, distribution, and integration to interactive processes, jointly constructing a theoretical framework of collective intelligence spanning “static structure to dynamic evolution.” Although these three mechanism models explain the emergence of collective intelligence from different perspectives, in real team collaboration practice, these three mechanisms are not isolated but rather an organic whole that intertwines and synergistically interacts. Examining them in combination, rather than separately, can greatly deepen our understanding of how collective intelligence “emerges” from individual interaction. Shared mental models represent the shared characteristics of team cognition, ensuring all members possess consistent understanding and expectations of goals, strategies, roles, and situations (Cannon-Bowers et al., 1993), providing a cognitive foundation and collaboration blueprint for effective team coordination and communication as a prerequisite; transactive memory systems represent the distributed characteristics of team cognition, serving as the core operational engine in dynamic interaction, ensuring teams can flexibly and efficiently utilize cognitive resources distributed across different members to respond to complex tasks; interactive team cognition models play the role of bridging static blueprints and dynamic engines, emphasizing the core process of communication and real-time interaction

(Cooke et al., 2013). Collective intelligence is born from the virtuous cycle and synergistic emergence among these three elements: shared cognitive foundations, efficient distributed division of labor, and continuous real-time interaction.

3 Measurement Paradigms for Collective Intelligence

Measurement paradigms for collective intelligence are essentially operational presentations of its mechanism models, with a deep theoretical mapping relationship between the two. Precise measurement of collective intelligence is both fundamental for optimizing organizational management and promoting human-AI collaboration, and an important direction for psychometrics to expand theoretical boundaries at the group level. Current research on collective intelligence measurement can be summarized into two paradigms, corresponding to different theoretical emphases of mechanism models. The **evaluative measurement paradigm** focuses on overall team performance, emphasizing comprehensive measurement of team effectiveness in collaborative tasks. This paradigm indirectly reflects the cognitive consistency emphasized by shared mental models and simultaneously embodies the team interaction effectiveness highlighted by interactive team cognition models through team performance on standardized assessment tasks. The **diagnostic measurement paradigm** focuses on analyzing the internal structure and mechanisms of team cognitive processes, typically employing multi-stage, decomposable experimental designs to separate individual contributions from team interaction effects, directly assessing the knowledge distribution and coordination characteristics emphasized by transactive memory systems. Simultaneously, this paradigm provides critical process analysis tools for interactive team cognition models by quantifying communication quality among members, transforming the theoretical proposition of “cognition as interaction” into observable and diagnosable empirical indicators. The choice of measurement paradigm depends on both researchers’ theoretical positions and specific research questions. The following sections systematically elaborate on the characteristics of these two measurement paradigms.

3.1 Evaluative Measurement Paradigm for Collective Intelligence

The evaluative measurement paradigm for collective intelligence focuses on measuring the level of collective intelligence. This approach uses the team as the unit of analysis and measures team-level comprehensive cognitive ability through standardized collaborative task performance. It emphasizes direct assessment of team final outputs and overall effectiveness, typically employing structured team tasks such as collective problem-solving, multi-round decision-making, or innovative ideation to elicit collaborative behaviors, and quantifies collective intelligence levels based on indicators such as task completion quality and efficiency.

The collective intelligence test developed by Woolley’s team represents a proto-

typical tool of this paradigm. Riedl et al. (2021) systematically measured collective intelligence through standardized tasks and statistical modeling methods. This study designed eight task types covering cognitive reasoning, collaborative integration, and coordination strategies, including brainstorming, matrix reasoning, memory tasks, Sudoku, typing tasks, and word decoding. Specific tasks and scoring methods are detailed in Table 2. Using the digital experimental environment of POGS (Platform for Online Group Studies), researchers achieved behavioral log data collection of member interaction trajectories, effectively ensuring measurement objectivity and replicability. At the data analysis level, Riedl and colleagues innovatively employed a two-stage structural equation model, conducting meta-analysis on covariance matrices from 22 independent studies to confirm the structural stability and measurement consistency of the *c*-factor across task contexts. The Collective Intelligence Score (CI Score) is converted to standardized Z-scores after extracting common factor loadings through principal component analysis, demonstrating significant explanatory power in predictive validity tests under novel task conditions. Additionally, based on team process data analysis, Riedl et al. (2021) constructed three major indicators to assess team interaction quality and performance: individual skill consistency (matching degree between member expertise and task requirements), strategy coordination (completeness of task path coverage), and interaction coordination quality (communication frequency).

The evaluative measurement paradigm directly assesses overall cognitive effectiveness by quantifying team performance across multidimensional standardized tasks. Such tasks can effectively elicit team collaborative behaviors and enable simultaneous standardized testing of multiple teams through online platforms while automatically collecting fine-grained process data (e.g., chat records). This high-throughput data collection approach significantly improves research efficiency, reduces costs, enables measurement and comparison of large team samples, and provides rich data foundations for subsequent statistical analyses. This paradigm is not limited to final performance scores but further utilizes collected fine-grained behavioral logs to construct indicators such as “individual skill consistency” and “strategy coordination,” thereby preliminarily parsing underlying interaction mechanisms while evaluating overall effectiveness. However, this paradigm also has inherent limitations. First, although it effectively assesses collective intelligence levels, it struggles to reveal the specific cognitive or behavioral mechanisms underlying results. Second, using team-level scores as output may mask unique contributions of individual members, making them unidentifiable through averaging effects. Finally, task design is particularly critical and complex, requiring careful consideration of the match between team size and task type/difficulty to ensure tasks can effectively stimulate team collaboration potential without compromising measurement validity due to excessive simplicity or complexity (Woolley et al., 2010).

3.2 Diagnostic Measurement Paradigm for Collective Intelligence

As research has deepened, the focus of collective intelligence measurement has gradually shifted from merely assessing collective intelligence levels to exploring key factors influencing these levels. The core task of the diagnostic measurement paradigm is to answer a critical question: whether poor team collaboration results from inadequate individual-level cognitive abilities or from structural or interactional problems in team collaboration processes. Answering this question helps more accurately diagnose weak links in team operations and provides targeted improvement directions for enhancing overall team performance.

The diagnostic measurement paradigm does not focus solely on final team outputs but rather achieves decomposition and diagnosis of the “individual cognitive input—team interaction process—team consensus output” chain through multi-stage data tracking, individual baseline establishment, and process proxy indicator construction, providing an operational analytical framework for understanding collective intelligence formation mechanisms. The ETS Tetralogue (Hao et al., 2015; Scoular & Care, 2019) represents a prototypical tool of this paradigm. Building on Zapata-Rivera et al. (2014), Hao et al. (2015) designed two simulation tasks: “Volcano single” and “Volcano CPS.” The difference between the two tasks is that the latter incorporates an embedded real-time communication interface, integrating authentic interpersonal collaboration elements while maintaining highly controlled experimental conditions. The task is designed as a standardized four-stage sequence: “independent response—collaborative discussion—answer revision—team submission,” systematically collecting individual responses and team dialogue data stored in structured logs. At the data analysis level, based on the PISA 2015 collaborative problem-solving competency framework, dialogue content is categorized into four types of social skills, with word count used as a proxy indicator for measuring collaboration engagement. In performance assessment, individual ability baselines are established by examining consistency between initial responses on independent and collaborative tasks for the same participants, ensuring they serve as effective covariates in subsequent modeling. Furthermore, performance improvement effects from collaboration are tested by comparing differences between revised and initial answers; team decision-making mechanisms are evaluated by comparing team final outputs with individual-level answers; and finally, a response change parameter (defined as the difference between the sum of two participants’ revised scores and the sum of their initial scores on the same item) is constructed to quantify overall cognitive gains from collaboration, thereby achieving systematic diagnosis of team interaction mechanisms and performance formation pathways.

The diagnostic measurement paradigm achieves a critical breakthrough in collective intelligence research from ability assessment to dynamic mechanism diagnosis through multi-stage experimental design. Its core value lies in precisely separating individual cognitive foundations (initial answer quality) from team collaboration effects (post-discussion answer improvement magnitude), thereby effec-

tively distinguishing whether poor team performance stems from member ability deficits or collaboration process failures (Hao et al., 2015), and consequently providing more targeted improvement directions for enhancing team effectiveness. Although this paradigm enhances control and comparability through highly structured task flows, it sacrifices some spontaneity and complexity of real team collaboration to some extent, resulting in reduced ecological validity. Second, relying on surface-level indicators such as word count to measure collaboration activity fails to capture deep cognitive processes such as perspective-taking and shared understanding construction (Guo Shaoyang & Yang Xiangdong, 2023). Additionally, static assumptions about individual abilities may not fully reflect dynamic learning and adaptive changes of team members during collaboration processes.

The evaluative and diagnostic measurement paradigms originate from different research objectives and thus carry distinct value functions. The evaluative paradigm provides a ruler for “knowing what is,” enabling effective measurement of collective intelligence levels; the diagnostic paradigm provides a lens for “knowing why,” profoundly revealing the collaborative processes and interaction mechanisms underlying collective intelligence. To deepen understanding of the applicable scenarios for these two measurement paradigms, we now present a horizontal comparison from perspectives including measurement focus, core assumptions, measurement tools, advantages and limitations, and applicable scenarios, as shown in Table 3 .

4 Discussion and Outlook

Collective intelligence, as a team-level cognitive capability, demonstrates its core value through synergistic effects that transcend simple aggregation of individuals via member collaborative interactions. In theoretical construction, shared mental models, transactive memory systems, and interactive team cognition models explain the cognitive foundations and effectiveness generation mechanisms of team collaboration from different perspectives, jointly constituting a multidimensional explanatory framework for collective intelligence—shared mental models elucidate the foundational role of cognitive alignment (Cannon-Bowers et al., 1993), transactive memory systems theory reveals optimization mechanisms of knowledge distribution (Wegner et al., 1991), while interactive team cognition models emphasize the dynamic characteristics of real-time interaction (Cooke et al., 2013). At the methodological level, collective intelligence measurement research has primarily formed two representative paradigms: the evaluative measurement paradigm (i.e., collective intelligence measurement components) focuses on overall team performance and interaction processes, while the diagnostic measurement paradigm (i.e., ETS Tetralogue) attempts to separate individual cognitive foundations from team collaboration effects to effectively distinguish root causes of poor team performance. Both paradigms have distinctive features but also obvious limitations: the evaluative paradigm efficiently quantifies overall effectiveness and collects process data through standardized tasks,

facilitating large-sample comparisons but offering insufficient revelation of deep mechanisms; the diagnostic paradigm precisely identifies team problem sources by separating individual abilities from collaboration effects through multi-stage designs but struggles to measure deep collaboration quality. This “evaluation-diagnosis” evolution in measurement paradigms reflects researchers’ deepening understanding of collective intelligence’ s essence. Based on the above theoretical and methodological developments and limitations, future research can deepen understanding and measurement of collective intelligence from three directions: first, in terms of measurement paradigm integration and optimization, innovative task design combined with multi-dimensional process data analysis can construct a comprehensive measurement framework for collective intelligence that balances ecological validity and mechanism diagnosis. Second, we should 致力于构建多模态动态评估体系, integrating behavioral, physiological, and neural data sources to achieve quantitative analysis of the entire team interaction process. Finally, for human-AI collaborative team research, we need to systematically explore its research paradigms and methodologies, focusing on frontier challenges such as cognitive division of labor, coordination mechanisms, and trust establishment between humans and intelligent agents.

Integration and Optimization of Measurement Paradigms: Achieving Dual Goals of “Accurate Measurement” and “Clear Diagnosis.” In measurement tool development, integrating the advantages of evaluative and diagnostic paradigms can follow evidence-centered design principles (Messick, 1992) to develop collective intelligence measurement tools. For instance, systematically designing information asymmetry tasks by differentially distributing complementary or partially conflicting information, goals, or knowledge among members in a controlled manner simulates information distribution states in real team decision-making, thereby effectively eliciting cooperation processes (Hesse et al., 2015). Such tasks break the condition of complete information sharing, forcing teams to integrate distributed cognitive resources through active communication, perspective articulation, critical questioning, and collective reasoning to reach consensus or decisions. In this process, researchers can not only evaluate team final performance but also finely capture key interaction sequences such as questioning, responding, consensus building, and conflict resolution, achieving effective observation and diagnosis of collaborative processes and deep cognitive mechanisms (Janssens et al., 2022; Yuan et al., 2019). In terms of psychometric models, specialized analytical tools suitable for team data need further development. Traditional models such as item response theory and cognitive diagnostic models have obvious limitations in team-level analysis due to difficulties in handling nested data structures and inter-member interaction dependencies. Recently, researchers have begun constructing team cognition diagnostic models to overcome these limitations. For example, the Team-DINA model can simultaneously assess team-level and individual-level cognitive attribute mastery and cognitive ability levels through analysis of team collaborative response data (Zhan et al., 2026); the Team-CDMs framework captures inter-member dependencies by jointly modeling individual cognitive attributes and “team-

work quality” reflecting social dimensions, thereby identifying team cognitive strengths and weaknesses and distinguishing whether poor performance stems from cognitive deficits or social problems (Zhan et al., 2025). These models provide feasible analytical pathways for refined diagnosis and mechanism analysis of team cognition. Meanwhile, increasingly mature large language models can provide solid technical foundations for team behavior analysis. This technology can not only reduce the workload of manual coding of team behaviors but also enable deep analysis of team internal communication, collaboration patterns, and their dynamic evolution. For example, natural language processing tools can be used for exploratory analysis, construct labeling, and construct-behavior association; text features generated by natural language processing can be used in regression models to study the impact of specific features on psychology or behavior (Feuerriegel et al., 2025).

Multimodal Dynamic Assessment System: Capturing Implicit Team Cognitive Processes. To overcome the limitation of traditional measurement relying solely on explicit behavioral data, future research can fully utilize emerging technologies to construct a “behavior-physiology-neural” trinity multimodal dynamic assessment system, providing new technical pathways and theoretical perspectives for deeply analyzing team cognitive processes. This comprehensive approach not only enables fine-grained decoding of team cognitive processes, revealing dynamic interaction mechanisms between individuals and groups, but also provides multidimensional validity evidence for measurement tool development. During testing, multiple types of multimodal data are often collected in real-time, including process data such as response accuracy and response time, as well as physiological signals, eye-tracking trajectories, and user interface interaction behaviors, to comprehensively capture examinees’ cognitive and behavioral characteristics (Liu Yaohui et al., 2022). Wang et al. (2025) analyzed intracranial electrophysiological signals in the amygdala and temporoparietal junction using time-frequency analysis methods, finding that these regions exhibited higher levels of inter-brain neural synchronization in the gamma band during cooperation establishment and maintenance states, with the degree of inter-brain neural synchronization predicting behavioral coordination levels between collaborators. Zhang et al. (2024) used functional near-infrared spectroscopy (fNIRS) hyperscanning technology to find that inter-brain synchronization and performance under collaborative conditions were superior to independent or competitive conditions in cooperative tasks. Lu et al. (2023) used fNIRS hyperscanning to reveal the neural inter-brain interaction basis underlying behavioral characteristics of high-creative groups during dynamic creative processes, finding that intra-brain and inter-brain neural coupling may be related to complex cognitive activities such as idea generation, idea evaluation, concept integration, and perspective-taking in high-creative groups during dynamic creative processes. Atweh et al. (2023) used eye-tracking technology based on Area of Interest Cross-Recurrence Quantification Analysis (CRQA) metrics (Laminarity: LAM, referring to the percentage of recurrent points forming vertical lines, indicating the proportion of time team members stay in the same area; Trapping Time: TT,

referring to the average time two trajectories stay in the same area, an indicator of team members' tendency to concentrate on certain screen regions) to understand collaboration quality among team members. Multimodal data processing involves complex key stages including data integration and fusion, preprocessing and feature extraction, and model training and evaluation. Specifically, regarding data integration and fusion, information from different sources needs to be integrated, such as ensuring temporal alignment of all data streams through specific software or programs (Klump et al., 2025). In preprocessing and feature extraction, data are preprocessed using methods such as noise reduction, artifact removal, and standardization, followed by feature extraction using methods like canonical correlation analysis and temporal alignment mechanisms to align and extract meaningful features from multimodal data (Hou et al., 2024). In model training and evaluation, multimodal machine learning or deep learning models are employed for joint modeling, with model performance, generalization capability, and robustness evaluated through cross-validation and external dataset testing. For instance, research has used multimodal deep learning models based on attention mechanism long short-term memory networks with physiological (blood volume, skin conductance, skin temperature) and behavioral (facial expressions, voice, keystrokes) data to identify human emotional states (Yang et al., 2023).

Collective Intelligence Research Based on Human-AI Collaboration: Exploring New Paradigms for “Human-AI Agent” Teams. With the deep development of artificial intelligence technology, human-AI collaboration will undoubtedly become the mainstream model for modern team work and intelligent system integration, with its core focus being the deep interaction and collaboration between human cognitive capabilities and AI (Cui & Yasseri, 2024). Brain-computer interfaces are recognized as core technologies for next-generation human-computer interaction and human-computer hybrid intelligence (Liu et al., 2025). Gupta et al. (2023) advocate using the Transaction Systems Model and instance-based learning theory to systematically design and understand future human-AI collaborative intelligent systems. In this context, how to effectively measure collective intelligence in human-AI collaborative contexts has become a key research issue. Based on existing literature, this issue can be systematically addressed at three levels: First, the dimensions of collective intelligence need to be redefined. While traditional teamwork emphasizes coordination and information integration among human members, in human-AI collaborative environments, this concept should be expanded to include new interaction mechanisms between humans and AI systems, including collaborative competencies such as trust establishment, dynamic role allocation, intention understanding, and capability complementarity. Second, appropriate measurement tasks should be carefully selected. Existing research indicates that human-AI collaboration effectiveness significantly depends on task types and relative AI performance (Vaccaro et al., 2024). Common experimental tasks include content generation (Li et al., 2024) and creative problem-solving (Sun et al., 2024). Notably, AI can already complete Raven's Progressive Matrices

based on visual imagery (Kunda, 2020), indicating such tasks are also suitable for evaluating cognitive synergy effectiveness in human-AI collaboration. Finally, a multi-level collaboration quality assessment system should be constructed, specifically covering three aspects: first, task performance indicators, i.e., evaluating team performance in quality, efficiency, and innovation through standardized tasks; second, interaction log analysis, i.e., analyzing behavioral indicators such as communication frequency, prompt length, and interaction rounds through AI system records (Li et al., 2024), with further identification of collaboration pattern characteristics (Wang Fancong et al., 2025); third, socio-emotional level measurement, i.e., evaluating team members' subjective collaboration experiences using tools such as trust scales and team climate questionnaires. However, this direction faces key challenges: first, cognitive biases between humans and AI may produce additive effects, such as groupthink polarization caused by algorithmic dependence (Peralta et al., 2021) and human-AI interaction feedback potentially amplifying human biases (Glickman & Sharot, 2025). Second, the dynamic evolution mechanism of bidirectional trust remains underdeveloped, requiring integration of human behavioral intention modeling and AI transparency enhancement technologies (Xie Yubin & Zhou Ronggang, 2025) to develop quantifiable trust assessment tools. This involves not only understanding how humans and AI establish mutual trust, maintain trust states, or lose trust over time during collaboration, but also designing methods and indicators that can accurately assess this bidirectional trust level. Third, the absence of an ethical framework leaves issues of responsibility attribution unresolved. Breakthroughs in this direction will drive collective intelligence research from an interpersonal collaboration paradigm to a human-AI collaboration paradigm, expand the intersection boundaries between cognitive science and artificial intelligence, and provide new perspectives for understanding "intelligence emergence" mechanisms.

References

- Bai Xinwen, Wang Erping. (2004). Research status of shared mental models. *Advances in Psychological Science*, 12(5), 791–799.
- Guo Shaoyang, Yang Xiangdong. (2023). Development and limitations of collaborative problem-solving competency assessment methods. *Global Education Outlook*, 52(6), 49–61.
- He Guibing, Yang Qiong. (2006). Measurement of shared mental models. *Chinese Journal of Ergonomics*, 12(4), 39–41. <https://doi.org/10.3969/j.issn.1006-8309.2006.04.013>
- Lin Xiaomin, Bai Xinwen, Lin Lin. (2014). Effects of similarity and accuracy of team mental models on team creativity. *Acta Psychologica Sinica*, 46(11), 1734–1747.
- Liu Yaohui, Xu Huiying, Chen Qipeng, Zhan Peida. (2022). Problem-solving competency measurement and data analysis methods based on process data. *Advances in Psychological Science*, 30(3), 522–538.

- Wang Fancong, Tang Xiaoyu, Yu Shengquan. (2025). Cognitive outsourcing based on generative artificial intelligence: Analysis of interaction behavior patterns and cognitive structure characteristics. *Acta Psychologica Sinica*, 57(6), 967–986.
- Xie Yubin, Zhou Ronggang. (2025). Bidirectional trust in human-AI relationships under new human-machine relationships. *Advances in Psychological Science*, 33(6), 916–932.
- Zhan Peida, Wang Zhimou, Chu Gaohong, Hao Ning. (2026). Team cognitive diagnosis modeling based on higher-order cognitive interactions among members. *Acta Psychologica Sinica*, 58(1), 166–179.
- Atweh, J., Hayek, J., & Riggs, S. (2023). Quantifying visual attention of teams during workload transitions using aoi-based cross-recurrence metrics. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 67(1), 1882–1887.
- Cannon-Bowers, J. A., Salas, E., & Converse, S. (1993). Shared mental models in expert team decision making. In N. J. Castellan, Jr. (Ed.), *Individual and group decision making: Current issues* (pp. 221–246). Lawrence Erlbaum Associates, Inc.
- Conradt, L., & Roper, T. (2003). Group decision-making animals. *Nature*, 421(6919), 155–158. <https://doi.org/10.1038/nature01294>
- Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive Science*, 37(2), 255–285. <https://doi.org/10.1111/cogs.12009>
- Cooke, N. J. (2015). Team cognition as interaction. *Current Directions in Psychological Science*, 24(6), 415–419. <https://doi.org/10.1177/0963721415602474>
- Cui, H., & Yasseri, T. (2024). AI-enhanced collective intelligence. *Patterns*, 5(11), <https://doi.org/10.1016/j.patter.2024.101074>
- Curşeu, P. L., Pluut, H., Boroş, S., & Meslec, N. (2015). The magic of collective emotional intelligence in learning groups: No guys needed for the spell!. *British Journal of Psychology*, 106(2), 217–234. <https://doi.org/10.1111/bjop.12075>
- Demir, M., Likens, A. D., Cooke, N. J., Amazeen, P. G., & McNeese, N. J. (2019). Team coordination and effectiveness in Human-Autonomy teaming. *Transactions on Human-Machine Systems*, 49(2), 150–159. <https://doi.org/10.1109/thms.2018.2877482>
- Feuerriegel, S., Maarouf, A., Bär, D., Geissler, D., Schweisthal, J., Pröllochs, N., ...Van Bavel, J. J. (2025). Using natural language processing to analyse text data in behavioural science. *Nature Reviews Psychology*, 4(2), 96–111. <https://doi.org/10.1038/s44159-024-00392-z>
- Galton, F. (1907). Vox Populi. *Nature*, 75, 450–451. <https://doi.org/10.1038/075450a0>

- Glickman, M., & Sharot, T. (2025). How Human-AI feedback loops alter human perceptual, emotional and social judgements. *Nature Human Behaviour*, 9(2), 345–359. <https://doi.org/10.1038/s41562-024-02077-2>
- Gorman, J. C., Grimm, D. A., Stevens, R. H., Galloway, T., Willemsen-Dunlap, A. M., & Halpin, D. J. (2020). Measuring real-time team cognition during team training. *Human Factors*, 62(5), 825–860. <https://doi.org/10.1177/0018720819852791>
- Gupta, P., & Woolley, A. W. (2021). Articulating the role of artificial intelligence in collective intelligence: A transactive systems framework. *Proceedings of the Human Factors and Ergonomics Society*, 65(1), 670–674. <https://doi.org/10.1177/1071181321651354c>
- Gupta, P., Nguyen, T. N., Gonzalez, C., & Woolley, A. W. (2023). Fostering collective intelligence in Human-AI collaboration: Laying the groundwork for COHUMAIN. *Topics in Cognitive Science*, 17(2), 189–216. <https://doi.org/10.1111/tops.12679>
- Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315–342). New York: Prentice-Hall.
- Hao, J., Liu, L., von Davier, A. A., & Kyllonen, P. (2015). Assessing collaborative problem solving with simulation based tasks. *International Conference on Computer Supported Collaborative Learning*.
- He, F., Pan, Y., Lin, Q., Miao, X., & Chen, Z. (2019). Collective intelligence: A taxonomy and survey. *IEEE Access*, 7, 170213–170225.
- Hesse, F., Care, E., Buder, J., Sassenberg, K., & Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin & E. Care (Eds.), *Assessment and teaching of 21st century skills* (pp. 37–56). Springer. https://doi.org/10.1007/978-94-017-9395-7_2
- Heylighen, F. (1999). Collective intelligence and its implementation on the web: Algorithms to develop a collective mental map. *Computational & Mathematical Organization Theory*, 5, 253–280. <https://doi.org/10.1023/A:1009690407292>
- Hou, K., Zhang, X., Yang, Y., Zhao, Q., Yuan, W., Zhou, Z., ...Hu, B. (2024). Emotion recognition from multimodal physiological signals via discriminative correlation fusion with a temporal alignment mechanism. *IEEE Transactions on Cybernetics*, 54(5), 3079–3092. <https://doi.org/10.1109/TCYB.2023.3320107>
- Janssens, M., Meslec, N., & Leenders, R. T. A. J. (2022). Collective intelligence in teams: Contextualizing collective intelligent behavior over time. *Frontiers in Psychology*, 13, 989572. <https://doi.org/10.3389/fpsyg.2022.989572>
- Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor? *Journal of Management*, 20(2), 403–437. [https://doi.org/10.1016/0149-2063\(94\)90021-3](https://doi.org/10.1016/0149-2063(94)90021-3)
- Klumpp, M., Embray, L., Heimburg, F., Dias, A. L. A., Simon, J., Groh, A., ...Both, M. (2025). Syntalos: A software for precise synchronization of simul-

- taneous multi-modal data acquisition and closed-loop interventions. *Nature Communications*, 16(1), 708. <https://doi.org/10.1038/s41467-025-56081-9>
- Krafft, P. M. (2018). A simple computational theory of general collective intelligence. *Topics in Cognitive Science*, 11(2), 374–392. <https://doi.org/10.1111/tops.12341>
- Kunda, M. (2020). AI, visual imagery, and a case study on the challenges posed by human intelligence tests. *Proceedings of the National Academy of Sciences of the United States of America*, 117(47), 29390–29397. <https://doi.org/10.1073/PNAS.1912335117>
- Kurvers, R. H. J. M., Wolf, M., Naguib, M., & Krause, J. (2015). Self-organized flexible leadership promotes collective intelligence in human groups. *Royal Society Open Science*, 2(12), 150222. <https://doi.org/10.1098/rsos.150222>
- Leitão, P., Queiroz, J., & Sakurada, L. (2022). Collective intelligence in self-organized industrial cyber-physical systems. *Electronics*, 11(19), 3213. <https://doi.org/10.3390/electronics11193213>
- Li, Z., Xu, Y., & Li, K. (2019). The influence factors of collective intelligence emergence in knowledge communities based on social network analysis. *International Journal of Intelligent Science*, 9(1), 23–43. <https://doi.org/10.4236/IJIS.2019.91002>
- Li, N., Zhou, H., & Mikel-Hong, K. (2024). Generative AI enhances team performance and reduces need for traditional teams. *arXiv*. <https://doi.org/10.48550/arxiv.2405.17924>
- Liu, Z., Mei, J., Tang, J., Xu M., Gao M., Wang K., ...Wu H. (2025). A memristor-based adaptive neuromorphic decoder for brain–computer interfaces. *Nature Electronics*, 8, 362–372 <https://doi.org/10.1038/s41928-025-01340-2>
- Lu, K., Gao, Z., Wang, X., Qiao, X., He, Y., Zhang, Y., & Hao, N. (2023). The hyper-brain neural couplings distinguishing high-creative group dynamics: An fNIRS hyperscanning study. *Cerebral Cortex*, 33(5), 1630–1642. <https://doi.org/10.1093/cercor/bhac161>
- Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process performance. *Journal of Applied Psychology*, 85(2), 273–283. <https://doi.org/10.1037/0021-9010.85.2.273>
- Messick, S. (1992). The interplay of evidence and consequences in the validation of performance assessments. *ETS Research Report Series*, 1992(1), i–42. <https://doi.org/10.1002/j.2333-8504.1992.tb01470.x>
- Mol, E., Khapova, S., & Elfring, T. (2015). Entrepreneurial team cognition: A review. *International Journal of Management Reviews*, 17(2), 232–255. <https://doi.org/10.1111/ijmr.12055>
- Narayanan, A., Korium, M., Carrillo, D., Hussain, H., Sena, A., Silva, P., ...Nardelli, P. (2022). Collective intelligence using 5g: Concepts, applications, and challenges in sociotechnical environments. *IEEE Access*, 10, 70394–70417.

- Palak, R., & Wojtkiewicz, K. (2021). The formal framework for collective systems. *Axioms*, 10(2), 91. <https://doi.org/10.3390/axioms10020091>
- Peralta, A. F., Neri, M., Kertész, J., & Iñiguez, G. (2021). Effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions. *Physical Review E*, 104(4), 044312. <https://doi.org/10.1103/PhysRevE.104.044312>
- Reitman, J. G., Harrison, J. L., Gorman, J. C., Lieber, R., & D' Mello, S. K. (2025). Communicative influence: A novel measure of team dynamics that integrates team cognition theory with collaborative problem solving assessment. *Journal of Educational Psychology*, 117(1), 134–151. <https://doi.org/10.1037/edu0000904>
- 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), e2005737118. <https://doi.org/10.1073/pnas.2005737118>
- Seely, T. D. (1995). *The wisdom of the hive: The social physiology of honey bee colonies*. Harvard University Press.
- Scoular, C., & Care, E. (2019). A generalized scoring process to measure collaborative problem solving in online environments. *Educational Assessment*, 24(3), 213–234. <https://doi.org/10.1080/10627197.2019.1615372>
- Siemon, D. (2022). Elaborating team roles for artificial intelligence-based teammates in Human-AI collaboration. *Group Decision and Negotiation*, 31(5), 871–912. <https://doi.org/10.1007/s10726-022-09792-z>
- Sun, L. N., Yu, Y., Yao, Y., Li, Y., Zhang, H., Xie, X., ...Stillwell, D. (2024). Large language models show both individual and collective creativity comparable to humans. <https://doi.org/10.48550/arxiv.2412.03151>
- 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 & Co.
- Vaccaro, M., Almaatouq, A., & Malone, T. (2024). When combinations of humans and AI are useful: A systematic review and meta-analysis. *Nature Human Behaviour*, 8(12), 2293–2303.
- Wang, D., Churchill, E. F., Maes, P., Fan, X., Shneiderman, B., Shi, Y., & Wang, Q. (2020). From Human-Human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. *Human Factors in Computing Systems*, 1–6. <https://doi.org/10.1145/3334480.3381069>
- Wang, J., Meng, F., Xu, C., Zhang, Y., Liang, K., Han, C., ...Ma, Y. (2025). Simultaneous intracranial recordings of interacting brains reveal neurocognitive dynamics of human cooperation. *Nature Neuroscience*, 28(1), 161–173.
- Wechsler, D. (1956). *Die Messung der Intelligenz Erwachsener. Textband zum Hamburg-Wechsler-Intelligenztest für Erwachsene (HAWIE); Deutsche Bear-*

beitung Anne von Hardesty, und Hans Lauber. [The measurement of adult intelligence]. Hans Huber.

Wegner, D. M., Erber, R., & Raymond, P. (1991). Transactive memory in close relationships. *Journal of Personality and Social Psychology*, 61(6), 923–929. <https://doi.org/10.1037//0022-3514.61.6.923>

Weller, J., Boyd, M., & Cumin, D. (2014). Teams, tribes and patient safety: Overcoming barriers to effective teamwork in healthcare. *Postgraduate Medical Journal*, 90(1061), 149–154. <https://doi.org/10.1136/postgradmedj-2012-131168>

Wheeler, W. M. (1922). Social life among the insects. *The Scientific Monthly*, 15(6), 527–541. <http://www.jstor.org/stable/6662>

Wildman, J. L., Salas, E., & Scott, C. P. (2014). Measuring cognition in teams: A cross-domain review. *Human Factors*, 56(5), 911–941. <https://doi.org/10.1177/0018720813515907>

Woolley, A. W., Aggarwal, I., & Malone, T. W. (2015). Collective intelligence and group performance. *Current Directions in Psychological Science*, 24(6), 420–424. <https://doi.org/10.1177/0963721415599543>

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., & Gupta, P. (2024). Understanding collective intelligence: Investigating the role of collective memory, attention, and reasoning processes. *Perspectives on Psychological Science*, 19(2), 344–354. <https://doi.org/10.1177/17456916231191534>

Wooldridge, A. R., Carayon, P., Hoonakker, P., Hose, B.-Z., Brazelton, T. B., Eithun, B., ...Gurses, A. P. (2019). Team cognition as a barrier and facilitator in care transitions: Implications for work system design. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 63(1), 648–652. <https://doi.org/10.1177/1071181319631303>

Yang, K., Wang, C., Gu, Y., Sarsenbayeva, Z., Tag, B., Dinger, T., ...Goncalves, J. (2023). Behavioral and physiological signals-based deep multimodal approach for mobile emotion recognition. *IEEE Transactions on Affective Computing*, 14(2), 1082–1097. <https://doi.org/10.1109/TAFFC.2021.3100868>

Yu, C., Chai, Y., & Liu, Y. (2018). Literature review on collective intelligence: A crowd science perspective. *International Journal of Crowd Science*, 2(1), 64–73. <https://doi.org/10.1108/IJCS-08-2017-0013>

Yuan, J. L., Xiao, Y., & Liu, H. Y. (2019). Assessment of collaborative problem solving based on process stream data: A new paradigm for extracting indicators and modeling data. *Frontiers in Psychology*, <https://doi.org/10.3389/fpsyg.2019.00369>

Zapata-Rivera, D., Jackson, T., Liu, L., Bertling, M., Vezzu, M., & Katz, I. R. (2014). Assessing science inquiry skills using dialogues. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Intelligent tutoring systems: ITS 2014* (pp. 625-626). Springer.

Zhan, P., Wang, Z., Chu, G., & Qiao, H. (2025). Teamwork cognitive diagnostic modeling. *Psychometrika*. Advance online publication, 90(4), 1319–1345. <https://doi.org/10.1017/psy.2025.10036>

Zhang, M., Yin, Z., Zhang, X., Zhang, H., Bao, M., & Xuan, B. (2024). Neural mechanisms distinguishing two types of cooperative problem-solving approaches: An fNIRS hyperscanning study. *NeuroImage*, 291, 120587.

Zheng, Y. (2012). Unlocking founding team prior shared experience: A transactive memory system perspective. *Journal of Business Venturing*, 27(5), 577–591.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv – Machine translation. Verify with original.