The International Journal of Advanced Manufacturing Technology (2026) 144:617–639
https://doi.org/10.1007/s00170-026-17947-4

ORIGINAL ARTICLE

Productivity assessment in multi-human multi-robot manufacturing teams: insights and proposal for novel KPIs

Aurora Sofia Di Matteo[1] · **Matteo Capponi**[1] · **Luca Mastrogiacomo**[1] · **Fiorenzo Franceschini**[1]

Received: 4 September 2025 / Accepted: 20 March 2026 / Published online: 29 March 2026

Abstract

Multi-Human Multi-Robot (MH-MR) teams composed of heterogeneous human and robotic agents are rapidly expanding across various sectors, including environmental disaster management, space exploration, search and rescue missions, healthcare, and manufacturing applications. While the integration of heterogeneous agents, humans and robots, introduces significant technical and operational challenges, an even greater complexity arises when evaluating their effectiveness and performance. As productivity of MH-MR teams remains a relatively unexplored field, existing assessment models and frameworks are limited and often fragmented. This article provides a focused analysis of productivity metrics, with primary reference to manufacturing contexts, critically evaluating these indicators in terms of their adaptability and limitations for MH-MR teams. Based on this, the paper proposes revised and novel indicators capable of reflecting the dynamics and coordination challenges typical of MH-MR teams. These findings aim to guide future research towards more comprehensive and effective assessment strategies. Examples from an explanatory assembly case study are presented to illustrate the practical application of these novel KPIs in real-world contexts.

Keywords Multi-human multi-robot teams · Human-robot collaboration · Manufacturing · Productivity metrics · Team assessment

List of Acronyms

MH-MR	Multi-Human Multi-Robot
HRT	Human-Robot Team
HRC	Human-Robot Collaboration
HRI	Human-Robot Interaction
BOM	Bill of Materials

✉ Fiorenzo Franceschini
fiorenzo.franceschini@polito.it

Aurora Sofia Di Matteo
aurora.dimatteo@polito.it

Matteo Capponi
matteo.capponi@polito.it

Luca Mastrogiacomo
luca.mastrogiacomo@polito.it

[1] DIGEP (Department of Management and Production Engineering), Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy

1 Introduction

The increasing integration of robots alongside human workers is shaping new paradigms of collaboration in industry, healthcare, services, and mission-critical domains including search and rescue operations, disaster response, and space exploration [2, 6, 13, 27, 33, 52]. In modern manufacturing environments, human workers now share tasks with collaborative robots (i.e., "cobots") on assembly lines [4, 32, 56]. Such a novel paradigm, commonly known as "Human-Robot Collaboration" (HRC), aims to combine human dexterity and problem-solving skills with robotic precision and repeatability to increase productivity while maintaining safety [4, 20, 39]. As real-world applications spread, interactions are shifting from simple dyads to Multi-Human Multi-Robot (MH-MR) teams, where humans and robots jointly pursue shared objectives [13]. The interest in MH-MR teams lies in their potential to combine complementary strengths of humans and robots: well-coordinated teams can improve productivity, robustness, and adaptability to changing conditions [45]. However, coordinating many agents can lead to emergent behaviours and interaction patterns not seen in one-to-one collaborations, resulting in increased cognitive demands and risk of operator fatigue, especially when control is not dynamically shared [13, 46]: human team members may need to divide their attention and tasks among multiple robots, while robots must interpret inputs from and respond to multiple humans in parallel. Adding members does not necessarily enhance performance: only careful assessments can reveal whether MH-MR teams operate efficiently, safely, and in a coordinated way [8, 41, 50, 51]. In this way, traditional metrics developed for simpler contexts may not be directly applicable to MH-MR teams where teamwork, group dynamics, and coordination become critical factors [28, 38, 41, 50]. Additionally, the Industry 5.0 paradigm increasingly emphasises human-centricity, resilience, and sustainability, requiring technical performance evaluations complemented by broader assessment frameworks, able to capture cognitive, organisational, and social aspects of MH-MR collaboration [7, 17, 37]. Recent scientific literature contributions have attempted to move in this direction. For example, notable progress has been made through the benchmarking model by Riedelbauch et al., [50], which proposes a preliminary integrated approach for analysing human-robot teams. This framework identifies various assessment dimensions, such as productivity, flexibility, job quality, and safety, and introduces a series of quantitative metrics to evaluate system performance. However, while these metrics are valuable, they have primarily been developed for dyadic human-robot teams, so their suitability for complex MH-MR teams is limited. This paper extends the work of Riedelbauch et al., [50] by focusing specifically on their "productivity" metrics and adapting them to the context of MH-MR teams in manufacturing assembly scenarios. Although the analysis is grounded in this domain, the proposed revisions represent a methodological contribution to support future adaptations to other MH-MR collaborative domains. In detail, this paper focuses on the following research question: what are the main limitations of the existing productivity metrics and how can they be revised for assessing productivity in MH-MR teams?

The article is organised as follows: Sect. 2 introduces the concept of MH-MR teams and the main critical factors; Sect. 3 analyses the existing productivity metrics selected for this study; Sect. 4 presents an assembly case study to verify the adaptability of these metrics to MH-MR teams; Sect. 5 discusses the proposed improvements and their implementation in the explanatory case study; Sect. 6 summarises the main results, limitations and discusses directions for future developments.

2 Literature review

As human-robot collaboration (HRC) moves beyond simple dyads, the need for clear definitions becomes increasingly important, particularly in complex manufacturing environments. Although the term "human-robot team" (HRT) is widely used, its meaning can vary significantly across studies and is often imprecise when applied to heterogeneous multi-agent systems. To provide a more solid basis for this work, the authors adopt the following definition of a MH-MR team: *"a Multi-Human Multi-Robot (MH-MR) team is a structured group composed of at least three heterogeneous agents, including both human and robotic members, working together towards a shared goal through continuous interaction, information exchange, effective coordination, clear task allocation, and dynamic collaboration between agents throughout the production process"*. Figure 1 conceptually illustrates this definition, highlighting heterogeneity, shared workspace interaction, and coordinated task execution among multiple agents.

This definition emphasises the team as an integrated system, in which interactions between agents, rather than simply coexistence, drive the overall performance. This is particularly important in manufacturing, where structured collaboration, task interdependence, and real-time coordination play a crucial role.

In fact, by adopting this refined definition, the paper avoids the ambiguity where the term "team" is used inconsistently, sometimes being applied to dyadic or coupled systems [14, 24, 26]. We also include cases that involve three agents in the team [1], two robots and one human, or vice versa, providing that they are actively cooperating to complete assigned missions and shared goals [13, 29]. However, adding multiple humans and robots to the same environment significantly increases complexity, raising challenges absent in dyadic systems and giving crucial meaning to the concept of teamwork [38]. MH-MR teams are inherently dynamic and heterogeneous: roles, task allocation, communication flows, and coordination patterns may continuously evolve throughout task execution [35, 58]. This adds uncertainty and non-linearity to the interaction, making it difficult to predict outcomes or apply static coordination models. Managing these interactions requires refined teaming strategies that support real-time adaptation, shared situational awareness, and cognitive load balancing across human and robotic agents [15, 21, 40, 55]. For these reasons, our definition includes the aspects of continuous interactions, information exchange, coordination, and dynamic collaboration [11].

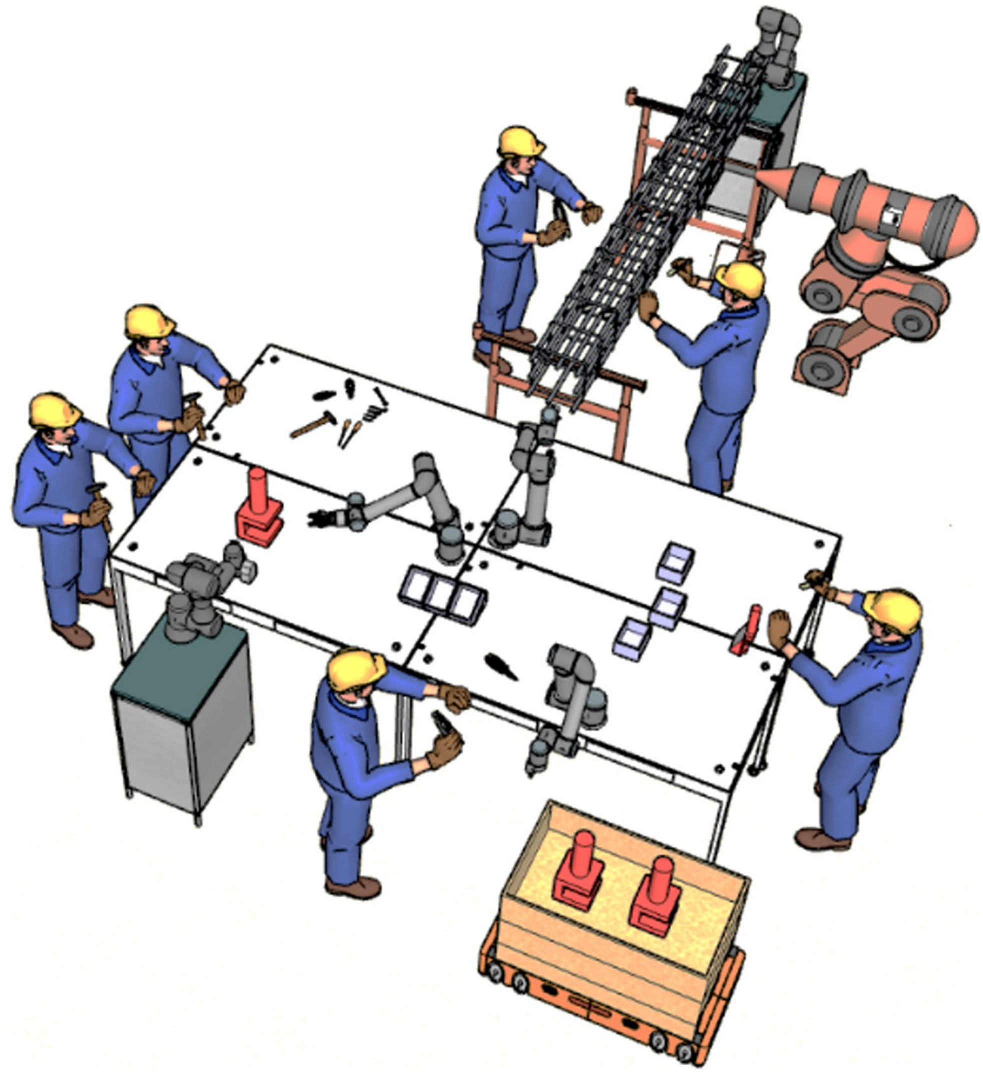


Fig. 1 An explanatory representation of a MH-MR team in a manufacturing scenario, showing heterogeneous human and robotic agents interacting within a shared workspace

Implementing such teams is a challenging activity. MH-MR teams are characterised by several structural factors that determine their success, and that should be taken into account in the performance assessment. Team size and team composition (i.e., the ratio of humans to robots in a team), if not properly designed, can undermine team cohesion, team performance, and member well-being [16]. Establishing cohesion among team members is necessary to build a seamless collaboration with their robotic counterparts [12]. Individual characteristics of each team member and the occurrence of external factors can impact on the outcome of the teaming [38]. Due to their intrinsic heterogeneity and evolving structure, MH-MR teams are subject to dynamic changes that may introduce additional variables, requiring careful coordination and control [58]. For example, tasks must be reassigned dynamically when agents' capabilities change because of fatigue, overload, failure, or evolving mission requirements [41, 59]. Human conditions, such as cognitive workload, performance, and emotions, may vary over time in response to internal or environmental factors [10, 29]. If not carefully managed, they can negatively affect the MH-MR team's overall performance. Identifying effective strategies to achieve better performance and ensure balanced workload distribution among all team members is a key challenge [29, 44]. Trust between team members, both human and robotic, plays a critical role in shaping the fluency of interaction, which in turn affects the quality of coordination and the team's overall effectiveness [14, 23, 25]. Communication is an important coordination mechanism that influences situational awareness and exchange of information among team members [19]. Negative effects of communication may increase workload, reduce performance, and interfere with the achievement of shared objectives [19, 44].

While MH-MR teams can offer significant advantages in terms of efficiency, flexibility, and coordination [29, 45], their effectiveness depends on the ability to manage complexity at multiple levels. This highlights the need for comprehensive and coherent approaches capable of capturing the full range of dynamics that characterise MH-MR teams [38, 49]. Models and metrics should support the understanding of how humans and robots interact and mutually adapt to each other. Traditional assessment methods developed for human-robot collaboration (HRC) and simpler human-robot interaction (HRI) scenarios may no longer be adequate or effective when applied to MH-MR team dynamics. Over the past two decades, researchers proposed numerous metrics and assessment methodologies, initially focused mainly on objective metrics [22, 43, 53, 54]. Later, research in HRC and HRI expanded to include subjective metrics related to the human operator's experience [5, 24, 31]. In parallel, other contributions addressed teamwork-specific metrics that reflect the quality of interaction and group coordination [3, 25, 28, 38]. Despite these advancements, these key indicators are originally developed for dyadic interactions and may not adequately reflect the challenges of larger and heterogeneous teams. In fact, the increased complexity of MH-MR teams amplifies coordination demands across physical, communicative, and cognitive dimensions, while exposing limitations in traditional performance metrics, that fail to account for new involved factors and teamwork aspects [38]. In particular, this paper focuses on the evaluation of productivity in MH-MR teams, by analysing selected existing metrics, and through their application to an explanatory case study, to support the development of effective assessment approaches.

To contextualise the maturity of research on mixed human-robot teams, a literature screening identified 330 journal articles published over the past decade. As shown in Fig. 2, the number of publications has steadily increased. However, only approximately 16% of these studies explicitly address industrial applications. This limited availability contrasts with the growing industrial adoption of collaborative robots. Market analyses project that the global collaborative robotics market will reach approximately USD 2.28 billion by 2026 and exceed USD 5.7 billion by 2031 (Mordor Intelligence, [42]). These trends highlight the need for structured productivity assessment approaches tailored to MH-MR manufacturing contexts.

3 Assessing MH-MR teams with a focus on productivity

Among the few assessment approaches presented in the literature, the work of Riedelbauch et al., [50] proposes a framework specifically designed for the assessment of human-robot teams. The framework explores different dimensions of analysis (productivity, flexibility, job quality, and safety), identifying key elements, and, in certain cases, recommending the use of targeted assessment metrics. Although it provides an initial structure for understanding the many critical aspects to be considered when implementing a MH-MR team, questions arise about the framework's ability to capture all the relevant dimensions and the effectiveness of the suggested metrics. Focusing on "Productivity", this paper critically analyses how the productivity metrics defined by Riedelbauch et al., [50] perform when applied to MH-MR teams. Table 1 reports a synthesis of the model proposed by Riedelbauch et al., [50].

The rest of this section is divided into seven subsections, each of which explains a productivity metric from the Riedelbauch's model.

3.1 Cooperative Speed-Up

The first analysed indicator is Cooperative Speed-Up ($S_{H/R}$), as follows:

Fig. 2 Annual number of journal publications on mixed human-robot teams (2015–2026)

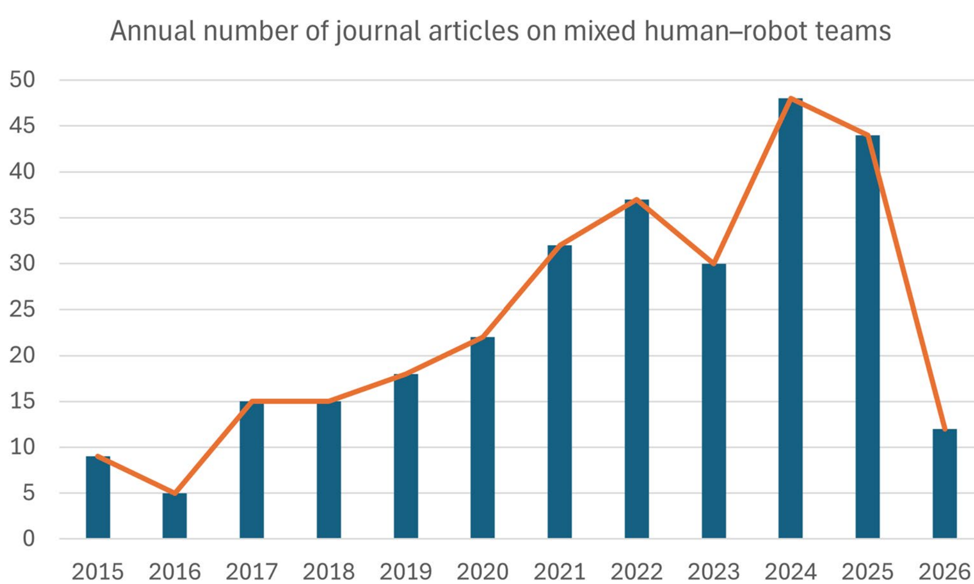


Table 1 Summary of the productivity metrics described by Riedelbauch et al., [50]

Metric	Definition
Cooperative Speed-Up	Ratio of the completion time of a human to that of a human-robot team. It measures process acceleration.
Relative Helpfulness	The percentage of time saved due to robot assistance.
Capability indicators	Quantitative measures to evaluate the suitability of a subtask for a human or a robot, based on their capabilities. Higher values indicate better alignment.
Robot Error Rate	Percentage of unsuccessful attempts by the robot.
Human/Robot Idle Time	The amount of time the human/robot is inactive during the process.
Concurrent activity	Duration where both human and robot are working simultaneously.
Robot Participation Rate	The percentage of subtasks performed by the robot compared to the total number of tasks.

$$S_{H/R} = \frac{D_H}{D_{H/R}} \tag{1}$$

where D_H is the time required for a human worker to manually complete a task (e.g., assembling a single product unit) and $D_{H/R}$ is the duration of completing the same task when partly automated by a human-robot team. These durations are derived using Methods-Time Measurement (MTM) tables [50] and do not capture execution variability, human learning effects, fatigue, or robot latency. The meaning of this metric is to compare the speed of execution between a human and a human-robot team and possibly see the reduction in time due to the benefits of working in a team. Cooperative Speed-Up domain is $S_{H/R} \in [0, +\infty)$. Its interpretation is described as follows:

- If $S_{H/R} > 1$, the human-robot team is faster than a single human.
- If $S_{H/R} = 1$, it means that D_H and $D_{H/R}$ are equal, which indicates that the human-robot team performs at the same speed as a single human.
- If $S_{H/R} < 1$, the human-robot team is slower than a single human.

3.2 Relative Helpfulness

The second metric defined is Relative Helpfulness (H_R), derived from the Cooperative Speed-Up $S_{H/R}$. It is defined as follows:

$$H_R = \frac{D_H - D_{H/R}}{D_H} = 1 - \frac{1}{S_{H/R}} \tag{2}$$

This metric quantifies the proportion of time saved when a human collaborates with a robot, rather than completing the task alone. Relative Helpfulness domain ranges from $-\infty$ to 1 ($H_R \in (-\infty, 1]$), although reaching a value of 1 is not realistic in practical scenarios, as it would imply that HRT completes the task in zero time. Its interpretation is described as follows:

- If $H_R \approx 1$, the robot reduces the cycle time to the minimum.
- If $H_R = 0$, it means that D_H and $D_{H/R}$ are equal, indicating that there is no improvement.
- If $H_R < 0$, the robot contributes to increasing the task duration due to inefficiencies or poor coordination.

Relative Helpfulness is a normalised measure of the robot's contribution expressed as a percentage of time saved. It is useful for assessing if the integration of the robot truly brings value.

3.3 Capability indicators

Capability Indicators assess how well the allocation of subtasks is in line with the specific characteristics of each team member (human and robot agents). As also said by Riedelbauch et al., [50], humans generally excel in dexterity and sensorimotor abilities, while robots are very good in precision or in repetitive tasks. Considering that, an efficient subtask allocation should reflect these differences and do not assign a difficult manipulation task to a robot and a repetitive positioning task to a human. In this case, there will certainly be suboptimal performance. So, these indicators quantify the suitability of each agent for a given subtask τ by assigning a real-valued score, denoted as $c_H(\tau)$ for humans and $c_R(\tau)$ for robots. Higher scores indicate better alignment between the agent's capabilities and the subtask requirements. The definition of $c_H(\tau)$ and $c_R(\tau)$ is obtained following different methods or approaches, like Ranz et al., [48], Lamon et al., [34] and Liau and Ryu [36]. Such capability indicators can be used in two ways:

- In pre-execution task allocation, as an optimisation criterion in static planning [30, 47];
- In post-execution, to evaluate the effectiveness of dynamic allocation decisions [47, 50].

3.4 Robot Error Rate

The Robot Error Rate (ε_R) is useful for understanding the impact of the human-robot team on productivity. It is defined as:

$$\varepsilon_R = 1\frac{N_R^{success}}{N_R^{attempt}} \tag{3}$$

where $N_R^{success}$ is the number of subtasks correctly completed by the robot and $N_R^{attempt}$ is the total number of attempts completed by the robot. The main task can be divided into different subtasks and then each of them can be evaluated to verify the correct execution. The Robot Error Rate domain is $\varepsilon_R \in [0, 1]$and its interpretation is described as follows:

- If $\varepsilon_R = 0$, it means that $N_R^{success}$ is equal to $N_R^{attempt}$ and that the robot makes no errors.
- If ε_R, it means that $N_R^{success} = 0$, suggesting that the robot always makes an error.

3.5 Human/Robot Idle time

Idle Time (D^{idle}) is an indicator of inefficiencies caused by temporal mismatches in coordination. It measures how long either the human (Human Idle Time, D_H^{idle}) or the robot (Robot Idle Time, D_R^{idle}) remains inactive during the execution of a task. This metric is useful for assessing the synchronisation between team members and identifying bottlenecks in collaboration. Idle times can be caused by task design, communication delays, or physical limitations. As these indicators represent how well production resources are utilised, their value should be minimised for optimal productivity. Therefore, they should be interpreted as follows:

- High Idle Times suggest poor coordination or task distribution.
- Low Idle Times indicate better synchronisation and continuous task engagement by both agents.

3.6 Concurrent activity

Concurrent activity (D^{coop}) evaluates the amount of time during which humans and robots work simultaneously in task execution. It is an important metric for understanding whether the team benefits from true collaboration, or if the process is mainly sequential. It can be interpreted as follows:

- A small deviation between the completion time of the HRT ($D_{H/R}$) and D^{coop} indicates successful parallelisation and efficient use of agent capacities.
- A large deviation between the completion time of the HRT ($D_{H/R}$) and D^{coop} implies that one agent is often idle, waiting for the other, leading to reduced team efficiency.

3.7 Robot Participation Rate

The Robot Participation Rate (P_R) represents how much the robot is involved in the process. Specifically, it is defined as:

$$P_R = \frac{N_R}{|T|} \tag{4}$$

where N_R is the number of subtasks handled by the robot, and $|T|$ is the overall number of subtasks in which the process is divided. Robot Participation Rate domain is $P_R \in$ [0,1] and its interpretation is as follows:

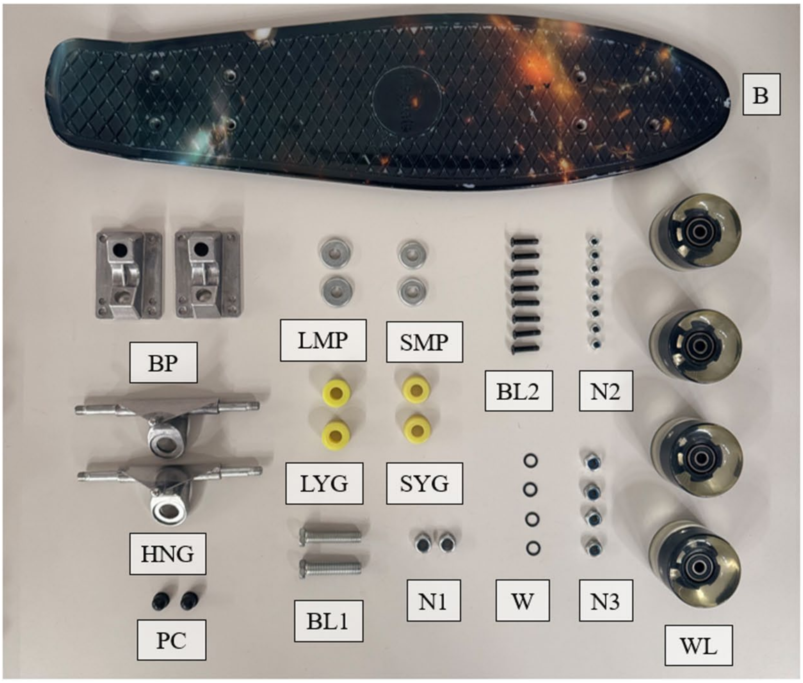
- If $P_R = 0$, the robot does not participate in the process at all ($N_R = 0$).
- If $P_R = 1$, only the robot performs the subtasks of the process (N_R is equal to $|T|$).

4 An explanatory assembly case study

To verify the applicability of these productivity metrics in the context of MH-MR teams, an explanatory assembly case study is considered. The task analysed is the assembly of a skateboard, whose Bill of Materials (BOM) is illustrated in Fig. 3. Two configurations are analysed:

- Manual configuration, with a single human agent.
- MH-MR configuration, with a team composed of two human operators and two collaborative robots (see Fig. 4).

The assembly process is divided into 3 macro steps, each of which must be done twice, once for the front and once for the rear of the skateboard: (A) assembly of the truck, (B) assembly of the wheels on the truck and (C) truck attachment on the board. These 3 macro steps are further broken down into subtasks, which are then allocated to team members based on their capabilities. For each configuration, the full task allocation is reported in Table 2, in which the last column "Time" represents the duration (in seconds) of each subtask of the process. In the Manual configuration, the subtasks S1 to S5 and S8 are repeated twice to assemble both



Code	Component	Quantity
B	Board	1
BP	Baseplate	2
HNG	Hanger	2
PC	Pin cup	2
LMP	Larger metal protector	2
SMP	Smallest metal protector	2
LYG	Larger yellow grommet	2
SYG	Smallest yellow grommet	2
BL	Bolt 1	2
N1	Nut 1	2
BL2	Bolt 2	8
N2	Nut 2	8
W	Washer	4
N3	Nut 3	4
WL	Wheel	4

Fig. 3 Bill of Materials (BOM) of the assembled skateboard

Fig. 4 MH-MR team in the assembly setup

sides of the skateboard and obtain the final product. Instead, in the MH-MR configuration the repetition of the tasks is divided between the two humans and the two robots.

The execution times reported in Table 2 are based on standard-time estimates. These values do not reflect learning dynamics, variability in execution conditions, potential failures or delays due to coordination.

To better understand the distribution of the subtasks among the two configurations, Gantt charts are reported in Figs. 5 and 6. These visual timelines show the sequence, the duration of each activity, and the eventual parallelism of subtasks, supporting the comprehension of the workflow organisation. From them, it is also possible to calculate the completion time of each assembly. Following the

Table 2 Skateboard's subtasks, related task allocation for both configurations with specific responsibilities and execution times

Assembly step	ID Subtasks	Subtasks	Configuration	Human 1	Human 2	Robot 1	Robot 2	Time (s)
A) Assembly of the truck	S1	Baseplate positioning	Manual	X				2 (for each truck)
			MH-MR			X	X	29
	S2	Truck assembly	Manual	X				35 (for each truck)
			MH-MR	X	X			25
B) Assembly of the wheels on the truck	S3	Wheel positioning	Manual	X				3 (for each wheel)
			MH-MR			X	X	8 (for each wheel)
	S4	Inserting wheels	Manual	X				3 (for each wheel)
			MH-MR	X	X			3 (for each wheel)
	S5	Wheel clamping	Manual	X				5 (for each wheel)
			MH-MR	X	X			5 (for each wheel)
C) Assembly of the truck on the board	S6	Base positioning	Manual	X				10
			MH-MR					-
	S7	Board positioning	Manual	X				5
			MH-MR			X	X	48
	S8	Truck positioning and bolting	Manual	X				50 (for each truck)
			MH-MR	X	X			40

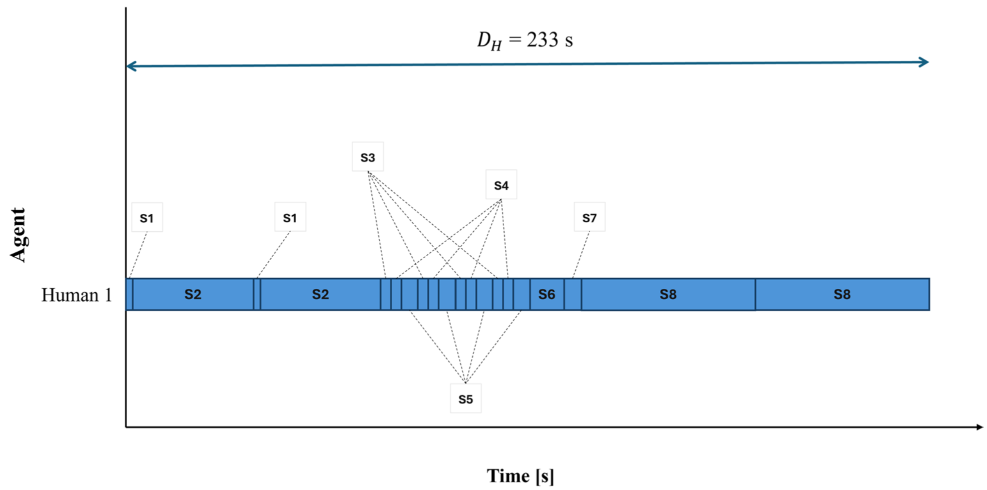


Fig. 5 Gantt chart of the assembly process performed by a single human operator (S1-S8 represent the process subtasks described in Table 2)

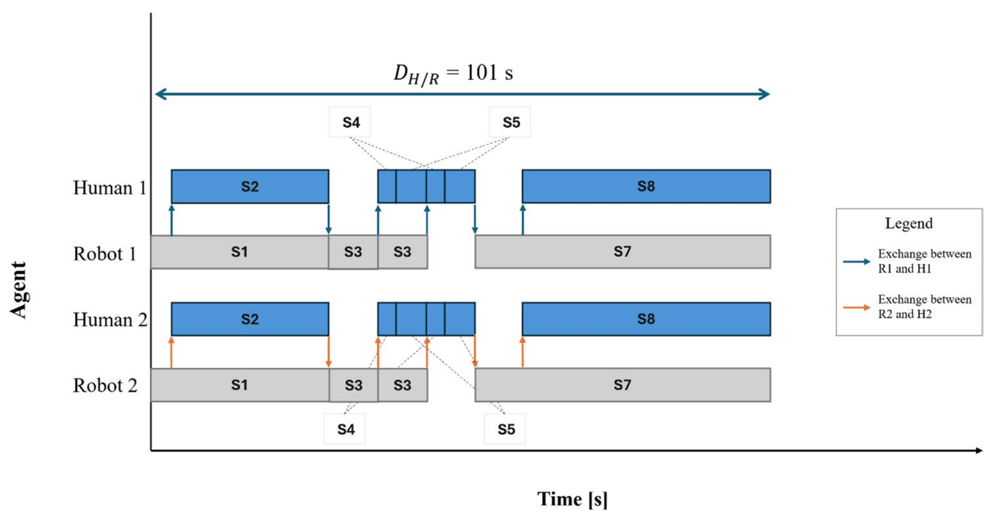


Fig. 6 Gantt chart of the assembly process performed by a MH-MR team (S1-S8 represent the process subtasks described in Table 2)

nomenclature of Riedelbauch et al., [50], the respective completion times of the single human and the MH-MR team are $D_H = 233\ s$ and $D_{H/R} = 101\ s$.

It is worth noting that this case study is an application designed to test how the selected productivity metrics behave in realistic MH-MR teams, in order to understand limitations and weaknesses of using them. The following subsections, from 4.1 to 4.8, critically analyse Riedelbauch's KPIs when applied to MH-MR teams for the assembly case study.

4.1 Cooperative Speed-Up in MH-MR team

In the case study, the manually executed assembly task required 233 s (D_H), whereas the same task performed by the MH-MR team was completed in 101 s ($D_{H/R}$). Based on these values, the Cooperative Speed-Up is calculated as:

$$S_{H/R} = \frac{D_H}{D_{H/R}} = \frac{233}{101} = 2.31 \tag{5}$$

According to its interpretation, this value indicates a reduction in the execution time during the assembly of the MH-MR team. The metric is able to reflect the temporal gain of adopting the MH-MR team, but there is no reference to the number or composition of the agents involved in the task. This can hide some critical aspects of team efficiency, such as resource utilisation, coordination, and exchange of information between members. Even if $S_{H/R} = 2.31$ indicates a speed-up, its value may overstate the benefit when the team size is ignored. Furthermore, the original formulation of Riedelbauch et al., [50] considers a dyadic collaboration between one human and one robot, making the metric less suitable for scenarios of multiple agents.

4.2 Relative Helpfulness in MH-MR team

Using the values from the case study, the Relative Helpfulness is calculated as follows:

$$H_R = \frac{D_H - D_{H/R}}{D_H} = 1 - \frac{1}{S_{H/R}} = 1 - \frac{1}{2.31} = 0.57 \tag{6}$$

The integration of an additional human and two robots into the MH-MR team resulted in a 57% gain in time efficiency. Since the Relative Helpfulness depends on the same input variables of Cooperative Speed-Up, they share the same structural limitations. Specifically, it does not consider the number and type of agents. In this case, quadrupling the number of agents (from one human to four agents) results in a 132 s (from 233 s to 101 s) gain in task execution time, without considering that the integration of new members can influence design and scalability decisions.

4.3 Capability indicators in MH-MR team

When applied to MH-MR contexts, Capability Indicators do not present significant limitations. In fact, they can support task allocation in heterogeneous teams by helping to match the characteristics of each subtask with the agent whose physical and cognitive capabilities are the best fit. These indicators can be computed using various approaches, ranging from structured quantitative methods to more qualitative assessments. In the presented case study, task allocation is based on the task assignment procedure defined by Bruno and Antonelli [9], that assesses four decision factors according to their absence (factor is equal to 0) or presence (factor is equal to 1) to decide if the subtask is more suited to a human or a robot:

- Weight (W), the load of the component. It is set to 1 for actions like lifting or holding over time, in which weight is a relevant factor.
- Displacement (Di), the distance of the component. It is set to 1 when tools and components are outside of the working area and are not easily reachable, so distance is a critical factor.
- Accuracy (A), the precision required in executing the subtask. It is set to 1 when the subtask involves high precision operations.
- Dexterity (De), the ability typical of humans to handle small or irregular objects. It is set to 1 when dexterity is needed to complete the subtask.

After the definition of these indicators, whose application to the case study is reported in Table 3, the assigned agent is deducted from the decision tree obtained by the authors Bruno and Antonelli [9] and shown in Fig. 7.

4.4 Robot Error Rate in MH-MR team

The Robot Error Rate (ε_R) is designed to evaluate the reliability of robotic agents by measuring the proportion of process failures relative to total attempts. In the case study,

Table 3 Capability indicators for the subtask assignment in the MH-MR configuration

ID Subtasks	Weight (W)	Displacement (Di)	Dexterity (De)	Accuracy (A)	Agent
S1	1	0	0	1	Robot
S2	0	1	1	0	Human
S3	1	0	0	1	Robot
S4	0	0	1	0	Human
S5	0	0	1	0	Human
S6	1	0	0	1	Robot
S7	0	0	1	0	Human

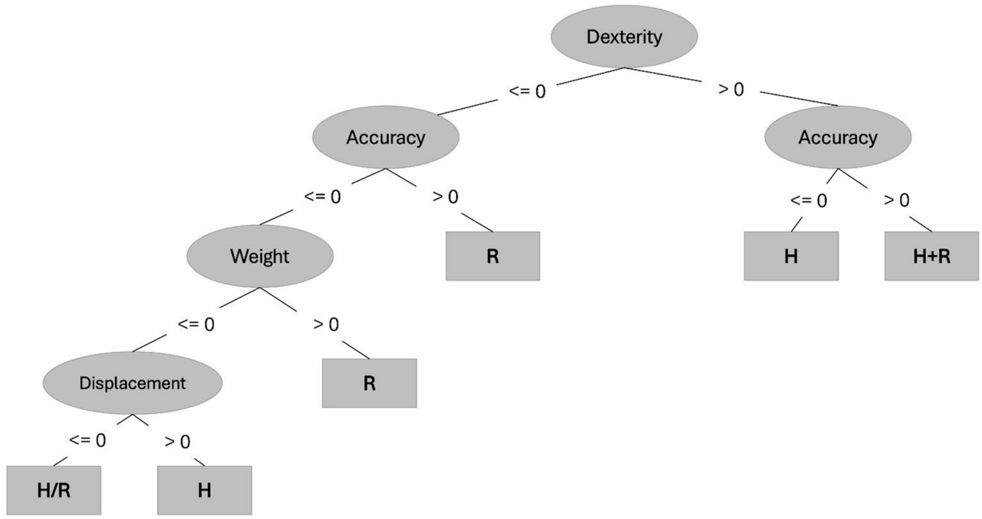


Fig. 7 Classification tree proposed by Bruno and Antonelli [9]

a total of 4 subtasks were allocated to the two robots (see Fig. 6), including: baseplate positioning, wheel positioning (repeated by each robot 2 times for taking one wheel at a time) and board positioning. All subtasks assigned to the robots were successfully completed, with no recorded errors. Therefore, the Robot Error Rate in this scenario is:

$$\varepsilon_R = 1 - \frac{N_R^{success}}{N_R^{attempt}} = 1 - \frac{4}{4} = 0 \tag{7}$$

This result indicates the highest level ($\varepsilon_R = 0$) of task reliability and coordination among the robotic agents, reflecting successful performance.

It is important to recognise that this case study is deterministic and does not incorporate stochastic failures or execution variability. In realistic MH-MR environments, non-zero error rates are to be expected due to the complexity of the task, sensor uncertainty, or mismatches in coordination. In these situations, unsuccessful attempts would increase ε_R and may require rework or human intervention, thereby affecting the overall duration of the task and productivity-related KPIs. For instance, if a robot fails to correctly place a wheel, a human operator must intervene to complete or repeat the subtask and the number of successful attempts ($N_R^{success}$) would decrease, while the total execution time (D_{MH-MR}) would increase, directly impacting productivity metrics and potentially altering configuration choices.

Since the metric was originally designed for single-robot systems [50], its interpretation in MH-MR contexts requires clarification regarding whether it should be evaluated at the individual level or through aggregation strategies across agents.

Table 4 Idle Times related to each configuration for each agent

Manual Assembly Idle Time	MH-MR Team Idle Times	
Human 1	Human 1 and 2	Robot 1 and 2
0 s	20 s (each)	8 s (each)

4.5 Human/Robot Idle Time in MH-MR team

The Idle Time metric measures the periods during which an agent is inactive, neither performing a task nor directly contributing to the workflow because of waiting for another agent. In its standard formulation, this metric is typically designed for systems involving a single human and a single robot. When applied to MH-MR teams, it's important to understand how to interpret idle times across multiple agents. Two primary approaches may be considered:

- Individual-based measurement, where the idle time of each agent is reported separately.
- Aggregated values, where average idle times per agent type or a global idle time per agent type can be computed.

In the case study, the results derived from the Gantt charts (Figs. 5 and 6) are reported in Table 4, obtained by subtracting from the completion time the sum of the periods during which each agent was occupied.

In this configuration, agents of the same type perform identical tasks in parallel, resulting in equal idle times and consistent conclusions between individual and aggregated reporting in MH-MR teams. However, in more heterogeneous MH-MR teams, where roles and task assignments differ among agents, relying only on aggregated idle time may hide important disparities in participation. In these contexts, reporting individual idle time becomes necessary for identifying unbalanced workloads or underutilised agents.

4.6 Concurrent activity in MH-MR team

Concurrent activity (D^{coop}) measures the time in which all the agents operate simultaneously during the process. The results in Table 5 were calculated by summing the overlapping time periods shown in the Gantt charts (Figs. 5 and 6), during which all the agents were active.

In the manual assembly, the operator is always working on the task, and the Concurrent activity is not calculated. In contrast, the MH-MR team achieves 73 s of full simultaneous activity among all four agents, out of a total task duration of 101 s.

While high Concurrent activity is generally interpreted as a sign of good coordination, particularly when it closely approaches the task duration, this assumption becomes problematic in MH-MR contexts. The more agents involved, the harder it is to achieve sustained parallel work across all team members. Therefore, although Concurrent activity provides a useful insight into teamwork dynamics, it strongly depends on team size, task structure, and interaction patterns. For this reason, its application in MH-MR settings may require complementary or refined indicators.

4.7 Robot Participation Rate in MH-MR team

The Robot Participation Rate (P_R) is intended to quantify how much the robot contributes to the overall process. In this case study, the total number of subtasks ($|T|$) is 20 (deduced by the Gantt chart in Fig. 6) and the number of subtasks performed by the robots (N_R) is 4 for each robot. This results in a Robot Participation Rate of:

$$P_R = \frac{N_R}{|T|} = \frac{8}{20} = 0.40 \tag{8}$$

This value suggests that robots are involved in 40% of the production process. As with other metrics, some considerations emerge. Firstly, it is necessary to determine if compute a single aggregated value (as done above since the two robots work simultaneously and do the same

Table 5 Concurrent activity computed for each configuration

Manual assembly D^{coop}	MH-MR team D^{coop}
Entire process performed by the single human	73 s

tasks) or separate values (in this case, we obtain a value of $P_R = \frac{4}{20} = 0.20$ for each robot). A separate P_R helps identify potential underutilisation or uneven distribution. Secondly, when subtasks are performed collaboratively, for example, by both a robot and a human simultaneously, or by two robots working on the same subtask (e.g., lifting a component together as the board in the case study), it is unclear whether to count them fully for each agent or apply partial weighting. Finally, the metric only considers the number of subtasks, ignoring their duration and assuming that all subtasks require similar execution times. As a result, a robot performing quick subtasks can have the same Robot Participation Rate as another robot executing the same number of longer, more demanding subtasks (see Fig. 8, where Robot 2 completes 4 short subtasks, reaching a higher P_R than Robot 1, which performs 3 longer subtasks, and it is involved for the entire execution time). These limitations suggest the revision of this metric to be adapted for MH-MR teams.

4.8 General considerations

This section summarises the findings emerged from applying each productivity metric to the case study. Although most metrics provide valuable insights, their use in more complex and heterogeneous team settings revealed several structural limitations.

In addition to metric-specific considerations, this case study reveals that the architecture of the team, its spatial configuration, its organisation between members and its dependence on how tasks are assigned to the agents play a critical role in determining productivity outcomes and are not explicitly captured by the analysed metrics. In the current setup, the MH-MR team is structured to allow parallel execution of certain subtasks. However, alternative

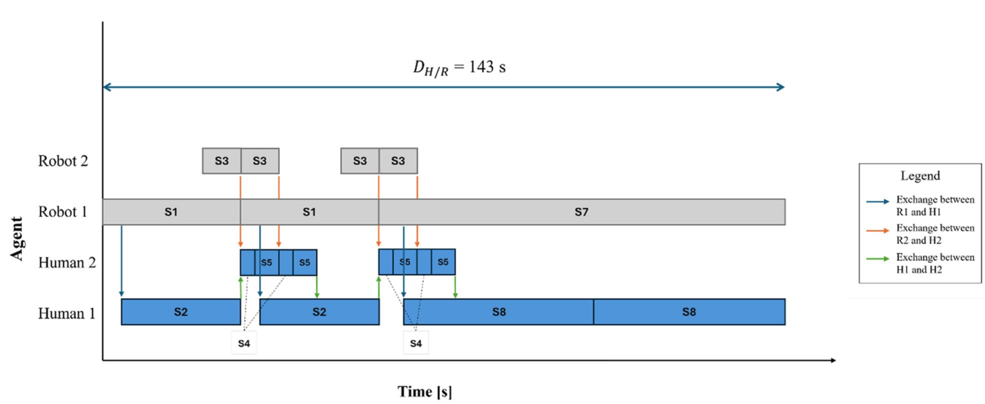


Fig. 8 Gantt chart of the assembly process performed by the MH-MR team in the second case (S1-S8 represent the process subtasks described in Table 6)

configurations (e.g., strictly sequential task flows) would lead to different productivity values. For example, Table 6 shows a more sequential allocation of the skateboard's subtasks among the same MH-MR team, composed of 2 humans and 2 robots. This alternative configuration affects the timing and the workflow of the assembly process (see Fig. 8 for the Gantt chart).

In this configuration, the total task completion time executed by the MH-MR team increases to 143 s, compared to the previous one (Fig. 6, cycle time = 101 s) achieved with the parallelised configuration. This increase is due to changes in task flow and coordination, not in task complexity or agent composition. As a result, all associated productivity metrics are affected.

From that, it is possible to highlight a key limitation of Riedelbauch's model productivity metrics: none of them explicitly account for team layout, how the subtasks are managed, or the type of interaction. In fact, simply rearranging how agents are organised, spatially or temporally, while keeping the task itself constant, can produce significantly different results. This underlines the need for more context-sensitive metrics that can incorporate team structure and task interdependencies.

Table 7 outlines each metric, its observed limitations in the MH-MR scenario, and possible improvements to extend their applicability in MH-MR teams discussed in Sect. 5.

5 Proposal for new MH-MR metrics

This section focuses on refining the productivity metrics, adapting them to better reflect the unique characteristics and complexities of MH-MR teams: critical factors such as agent heterogeneity, subtask allocation, team size, and interactions must be included [18, 57]. Each refined metric is defined in the relative subsection (from Sect. 5.1 to 5.7) and applied to the MH-MR case study. Section 5.8 then provides a general discussion of the integration and applicability of the proposed KPI framework.

5.1 Collaboration Efficiency (Revision of Cooperative Speed-Up)

The original metric of $S_{H/R}$, defined to assess dyadic human-robot collaborations through task completion time [50], shows limited effectiveness in MH-MR teams, composed of multiple heterogeneous agents (see Sect. 4.1). To address these limitations, two modifications have been introduced. The first adjustment removes the metric's insensitivity to team size and resource utilisation. Cooperative Speed-Up only considers efficiency in terms of task execution time, ignoring how many agents (human or robotic) were involved. The revised metric incorporates the total number of agents, enabling comparison between human only and MH-MR teams.

The second modification is based on the observation that the metric, even when adjusted for team size, does not distinguish the relative impact of different agent types on performance. In this regard, the new formulation introduces weight factors (α_h for human agents and α_r for robot agents), promoting balanced and meaningful task distribution within the team. In light of these changes, the metric has also been renamed Collaboration Efficiency, to reflect its broader evaluative scope, emphasising that efficiency comes from the quality of collaboration itself, including time, effort, and resource management. Collaboration Efficiency metric is defined as:

$$CE_{MH-MR} = \frac{n_{hM}}{\alpha_h * n_{hMH-MR} + \alpha_r * n_{rMH-MR}} * \frac{D_M}{D_{MH-MR}} \tag{9}$$

Where:

- D_M is the duration of the purely manual assembly process done by the human team;
- D_{MH-MR} is the duration of the MH-MR team assembly process;
- n_{hM} is the number of humans involved in the manual assembly (set equal to 1);
- n_{hMH-MR} is the number of humans in the MH-MR team;

Table 6 Skateboard's subtasks and related task allocation for the second MH-MR team configuration

Assembly step	ID Subtasks	Subtasks	Team	Human 1	Human 2	Robot 1	Robot 2	Time (s)
A) Assembly of the truck (done twice)	S1	Baseplate positioning	MH-MR			X		29
	S2	Truck assembly	MH-MR	X				25
B) Assembly of the wheels on the truck (done twice for each truck)	S3	Wheel positioning	MH-MR				X	8 (for each wheel)
	S4	Inserting wheels	MH-MR		X			3 (for each wheel)
	S5	Wheel clamping	MH-MR		X			5 (for each wheel)
C) Assembly of the truck on the board (done for each truck)	S7	Board positioning	MH-MR			X		85
	S8	Truck positioning and bolting	MH-MR	X				40

Table 7 Limitations and suggested improvements of productivity metrics for their applicability in MH-MR teams

Metric	Identified limitations	Proposed improvements
Cooperative Speed-up	− Does not account for the number and type of agents involved in task execution. − Same result for different team compositions if completion times are equal.	− Introduce number and type of agents as weighting factors. − Incorporate inter-agent dynamics to capture how task allocation and coordination affect performance.
Relative Helpfulness	− Same limitations as Cooperative Speed-Up. − Does not isolate the specific contribution of each robot.	− Refine based on enhancements to Cooperative Speed-up, including agent-specific contributions.
Capability indicators	− Its interpretation depends on the strategy used to match subtasks with agents.	− Use a flexible approach, choosing between qualitative or quantitative techniques based on which best fits the context.
Robot Error Rate	− Originally designed for single robot systems. − Lacks precision and detail when applied to MH-MR teams, where robots perform in different ways.	− Compute error rate per robot for the individual contribution or compute a global error rate to assess robotic reliability. − Extend the computation of Error Rates to human operators. − Define a Team Error Rate to assess overall system robustness.
Human/ Robot Idle Time	− May obscure participation disparities in agent utilisation.	− Clarify the analysis goal and its interpretation: report the individual values of each agent to understand the individual contribution or aggregate per type of agent for a broader overview.
Concurrent activity	− Difficult to achieve high values of concurrent activity with larger teams. − Not a direct indicator of productivity.	− Calculate Concurrent activity within subgroups (if applicable). − Use it in combination with other indicators to represent team collaboration.
Robot Participation Rate	− Only counts number of subtasks, ignoring task duration or difficulty. − Unclear how to handle shared tasks or multiple robots.	− Compute participation rates for each robot separately. − Aggregate using task-specific weighting to reflect shared efforts. − Consider a Robot Participation Rate based on execution times. − Extend the computation of Participation indicators to human operators. − Aggregate individual participation values to compute a Team Utilisation Indicator reflecting average resource engagement.

– $n_{r\,MH-MR}$ is the number of robots in the MH-MR team;
– α_h is the relative contribution of humans to the process (number of subtasks done by humans/total number of subtasks);
– α_r is the relative contribution of robots to the process (number of subtasks done by robots/total number of subtasks).

The factors α_h and α_r are derived from task allocation, but alternative weighting strategies might be adopted depending on the assessment goals and available data, such as basing the weights on execution times, cognitive workload, etc. Additionally, the choice of setting $n_{h\,M}$ equal to 1 is made to maintain consistency with the original indicator, which is recovered when both $n_{h\,MH-MR}$ and $n_{r\,MH-MR}$ are equal to 1. The domain of the revised indicator remains consistent with the original formulation $(CE_{MH-MR} \in [0, +\infty))$. In particular:

- If $CE_{MH-MR} > 1$, the MH-MR team is more efficient than the human team.
- If $CE_{MH-MR} = 1$, the MH-MR team obtains the same result of the human team.
- If $CE_{MH-MR} < 1$, the MH-MR team is less efficient than the human team.

A higher score indicates better collaborative effectiveness of the MH-MR team compared to the human team, considering the distinct contributions of human and robotic agents. Conversely, lower values may indicate inefficient allocation of resources, poor coordination, or unbalanced workload. Simply increasing the number of agents does not guarantee greater efficiency, especially if their coordination or performance is suboptimal. Using Eq. 10, the Collaboration

Efficiency of the MH-MR team compared to the single human worker is calculated as:

$$CE_{MH-MR} = \frac{n_{h_M}}{\alpha_h * n_{h_{MH-MR}} + \alpha_r * n_{r_{MH-MR}}} * \frac{D_M}{D_{MH-MR}} = \frac{1}{\frac{12}{20}*2+\frac{8}{20}*2} * \frac{233}{101} = 1.15 \quad (10)$$

The result indicates that the MH-MR team is approximately 15% more efficient than the manual configuration. The original Cooperative Speed-Up of 2.31 reflects a 57% reduction in completion time, but this value considers only task duration and not the number of agents involved. The Collaboration Efficiency normalises the time gain by accounting for resource utilisation. Therefore, the advantage gained through faster execution is partly offset by the increased number of agents deployed, resulting in a more modest overall improvement. In practical terms, values of Collaboration Efficiency below "1" indicate that the additional resources do not compensate for their cost in terms of productivity, whereas values slightly above "1" (e.g., 1.0–1.1.0.1) suggest marginal gains. Higher values indicate progressively more beneficial collaboration, although appropriate thresholds should be calibrated according to the specific operational and economic context. Meaningful thresholds should be derived from empirical studies across real MH-MR teams, taking into account operational costs, team composition, and task variability.

In the current formulation, the MH-MR team is compared to a single human reference. However, this metric can also be evaluated against alternative reference manual baselines performing the same assembly. For this purpose, a two-human configuration is defined, as detailed in Table 8, where task allocation and responsibilities are redistributed to maximise parallel execution while respecting task dependencies.

Based on the task durations reported from Table 8, the resulting duration of the task $D_{2H} = D_M$ is equal to 124 s and CE_{MH-MR} can be recomputed as:

$$CE_{MH-MR\,|\,2H} = \frac{n_{h\,M}}{\alpha_{\,h} * n_{h\,MH-MR} + \alpha_{\,r} * n_{r\,MH-MR}} * \frac{D_M}{D_{MH-MR}} = \frac{2}{\frac{12}{20}*2+\frac{8}{20}*2} * \frac{124}{101} = 1.23 \quad (11)$$

The value $CE_{MH-MR\,|\,2H} = 1.23$ indicates that the MH-MR configuration is more efficient than a parallelised two-human baseline. Compared to the value $CE_{MH-MR} = 1.15$ obtained when using a single-human baseline, the result shows that the MH-MR configuration's efficiency advantage persists and slightly increases, even compared to a stronger manual alternative. Although the number of human resources in the baseline doubles, the completion time does not decrease proportionally due to sequential dependencies and coordination constraints within the manual configuration, which lead to a higher Collaboration Efficiency value. Consequently, the MH-MR team achieves a comparatively lower completion time relative to the number of resources deployed, which reinforces the idea that the metric captures effective resource utilisation rather than absolute speed alone.

In this way, the new metric gives a more balanced and realistic assessment of MH-MR team performance, supporting more informed decisions.

5.2 MH-MR Relative Helpfulness

The original Relative Helpfulness quantifies the benefit gained in terms of time, by introducing a robot working with a human, but inherits the limitations of Cooperative Speed-Up (see Sect. 4.2). To address these issues, we introduce a revised version, named MH-MR Relative Helpfulness (H_{RMH-MR}), which builds on the Collaboration Efficiency metric presented in Sect. 5.1, and defined as:

$$H_{RMH-MR} = 1 - \frac{1}{CE_{MH-MR}} \quad (12)$$

This metric retains the conceptual structure of the original formulation but now is based on the Collaboration Efficiency. Its domain is $H_R \in (-\infty\,, 1]$ and maintains the original interpretation:

- If $H_{RMH-MR} \approx 1$, the MH-MR team is more helpful than the human team in both execution time and resource utilisation.

Table 8 Task allocation for the two-human manual baseline in the skateboard assembly case study

ID Subtasks	Subtasks	Configuration	Human 1	Human 2	Time (s)
S1	Baseplate positioning	Manual −2H	X	X	2
S2	Truck assembly	Manual −2H	X	X	35
S3	Wheel positioning	Manual −2H	X	X	3 (twice per human)
S4	Inserting wheels	Manual −2H	X	X	3 (twice per human)
S5	Wheel clamping	Manual −2H	X	X	5 (twice per human)
S6	Base positioning	Manual −2H	X		10
S7	Board positioning	Manual −2H		X	5
S8	Truck positioning and bolting	Manual −2H	X	X	50

Table 9 Examples of capability indicators for two subtask assignments in the MH-MR configuration of Table 6

ID Subtasks	Weight (W)	Displacement (Di)	Dexterity (De)	Accuracy (A)	Agent	Assigned agent
S1	1	0	0	1	Robot 1	Robot 1
S1	1	1	0	1	Robot 2	
S2	0	0	1	0	Human 1	Human 1
S2	0	1	1	0	Human 2	

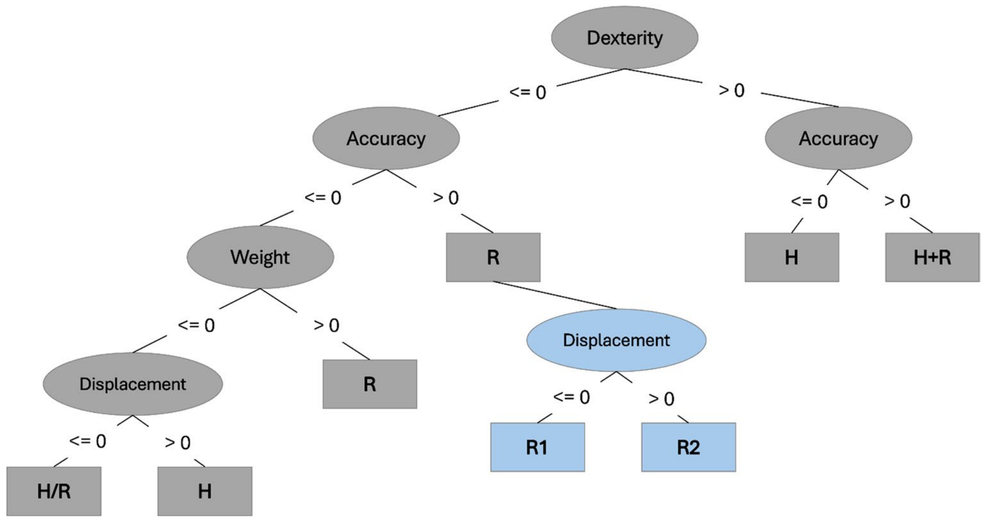


Fig. 9 Extension of the classification tree proposed by Bruno and Antonelli [9]

- If H_{RMH-MR}= 0, the MH-MR team performs equivalently to the human team; there is no improvement in efficiency or use of resources.
- If H_{RMH-MR}< 0, the MH-MR team is less effective compared to the human team.

High values of H_{RMH-MR} mean an effective integration of robots into the team. Instead, negative values indicate that involving robotic agents may be unjustified. MH-MR Relative Helpfulness considers team composition, including the number and type of agents in the team, facilitating comparisons between teams with different structures. Furthermore, through the weighting factors, α_h and α_r, it reflects the differentiated impact of humans and robots. Moreover, the metric considers both time and resource efficiency, moving beyond the simplified interpretation of speed gains alone.

MH-MR Relative Helpfulness applied to the case study can be calculated using Eq. (13):

$$H_{RMH-MR} = 1 - \frac{1}{CE_{MH-MR}} = 1 - \frac{1}{1.15} = 0.13 \tag{13}$$

This result indicates a 13% improvement in overall efficiency compared to manual assembly, accounting for time savings and team composition. Although lower than the 57% increase suggested by the original Relative Helpfulness, based solely on time reduction, it offers a more realistic view of whether the integration of more agents truly adds value in terms of productivity and scalability.

5.3 MH-MR capability indicators

In MH-MR contexts, defining capability indicators requires consideration on which type of agent is more appropriate for the execution of certain subtasks, as discussed in Sect. 4.3, but also on how choosing between agents of the same type. In the case study described, each human performs the same activities of the other, as well as robots do. Changing the architecture and the organisation of the task following a more sequential path as described in Table 6, we need also to decide the roles between agents of the same type. For that reason, the task assignment procedure executed in Sect. 4.3 should be extended and repeated for single agent, as illustrated in Table 9 for two subtasks. At this point, the classification tree shown in Sect. 4.3 has been expanded with an additional “Displacement” node (highlighted in light blue in Fig. 9), used to decide, for example, between Robot 1 or 2 for subtask 1. In this case, as Robot 2 is positioned outside the working area, distance becomes a critical factor (Di = 1), making Robot 1 the more suitable choice. Similar “Displacement” nodes should be added to the other responsibility rectangles to decide between the agents. However, they have been omitted from Fig. 9 for clarity.

This extension goes beyond simple task tagging as “human” or “robot”. In a basic classification, once a subtask is assigned to an agent type, the allocation process stops there. By contrast, the proposed MH-MR capability indicators introduce a hierarchical decision structure. After identifying the suitable agent type, additional contextual criteria determine the specific agent responsible for execution. For instance, if Subtask S1 requires robotic precision, a basic approach would assign it to a robot in general. However, in the extended structure, the displacement criterion differentiates between Robot 1 and Robot 2, selecting the agent best positioned to perform the task. This differentiation within types is particularly relevant in MH-MR configurations, where the effectiveness of tasks depends on contextual and organisational factors, and multiple agents of the same type coexist. Selecting the most appropriate agent in MH-MR teams should include considering individual capabilities, but also spatial proximity to the task location, ergonomic constraints, visibility, and freedom of movement, which are characteristics of each specific agent.

5.4 Error Rate indicators in MH-MR teams

As mentioned in Sect. 4.4, it is important to define the evaluation perspective to follow. Once the individual Robot Error Rate is computed, several strategies can be adopted to calculate an aggregate MH-MR Robot Error Rate, some of which are outlined below:

1. Mean Robot Error Rate: the average of all the individual rates, to have an overall indication of team performance.
2. Maximum Robot Error Rate: the highest individual error rate, to identify the weakest performance between robots and which would be critical.

Computing a single global error rate following certain strategies, such as the Mean Robot Error Rate, might obscure discrepancies between robots: one robot could perform perfectly while another frequently fails, yet the combined metric would not reflect this disparity. However, this global approach gives a broader overview of the robotic performance.

This formulation of individual Robot Error Rates can be extended to include individual Human Error Rates, focusing on the subtasks performed by human operators. The same aggregation strategies described above (e.g., Mean and Maximum Error Rate) can then be applied to calculate an aggregate MH-MR Human Error Rate. Moreover, a Team Error Rate, that accounts for all unsuccessful subtask attempts independently of the executing agent, can be computed in order to provide an overall indication of team robustness and its success.

5.5 MH-MR Human/Robot Idle time

As with the previous metric, it is possible to take a global perspective of the type of agent involved in the MH-MR team or conduct an individual analysis to determine each agent's resource usage. Section 4.5 reports the individual values of each agent. In this section, we will define ways to calculate MH-MR Human and Robot Idle Time. As these can be obtained in the same way, some alternatives for computing the MH-MR Human Idle Time are described below:

1. Mean Human Idle Time: the average of all the individual human idle times, to have an overall indication of the human agents.
2. Maximum Human Idle Time: the highest individual value, to identify the less involved agent between humans.

The MH-MR Human Idle Time for the first configuration remains the same, despite the adoption of different strategies, due to the symmetry of task execution. However, the two alternatives 1 and 2 can be applied to the second MH-MR team configuration ($n_{h\,MH-MR} = 2$). The individual values of each human can be deducted from the Gantt chart in Fig. 8 ($D^{idle}_{H1} = 13\ s$ and $D^{idle}_{H2} = 111\ s$). The results of MH-MR Human Idle Time are reported in Table 10.

The choice of how to aggregate MH-MR Human Idle Time depends on the specific goals of the analysis. Using the right strategy ensures that the metric aligns with the intended evaluation perspective. In the proposed MH-MR framework, Idle Time is not an isolated descriptive metric but a structural component of the participation indicators. In fact, Participation Indicators (Sect. 5.7) are directly derived from the complement of individual idle time with respect to total process duration. Consequently, variations in idle time modify participation values and influence the so named Team Utilisation Indicator (see Sect. 5.7).

5.6 MH-MR Concurrent activity

The Concurrent activity metric is a way of measuring the extent to which agents operate in parallel during task execution. However, as the size of the MH-MR team increases,

Table 10 Example of MH-MR Human Idle Time aggregation using two different strategies

MH-MR Human Idle time	Result
Mean Human Idle Time	$Mean_{MH-MR}\ D^{idle}_{H} = \frac{D^{idle}_{H1} + D^{idle}_{H2}}{n_{h\,MH-MR}} = \frac{13+111}{2} = 62\ s$
Maximum Human Idle Time	$Max_{MH-MR}\ D^{idle}_{H} = \max\left(D^{idle}_{H1};\ D^{idle}_{H2}\right) = \max\ (13;111) = 111\ s$

it becomes progressively less likely that all agents (both human and robotic) will be active simultaneously throughout the process. In larger teams, concurrent activity tends to be localised within specific subgroups or task phases. It is important to contextualise the interpretation of Concurrent activity with respect to the team structure, including its composition, the task allocation strategy, and the underlying collaborative architecture. The expected level of concurrency is greatly influenced by whether agents are arranged hierarchically, work in parallel streams, or follow a sequential handover.

Concurrent activity captures a dimension that is not reducible to participation or efficiency indicators. While Idle Time and Participation Indicators quantify individual engagement, Concurrent activity measures the extent to which all agents are simultaneously active, reflecting the team's structural coordination pattern. Two configurations may exhibit similar participation or efficiency values while differing substantially in synchronisation. Therefore, Concurrent activity is retained as an independent metric to preserve information about collaborative structure that would otherwise be lost in aggregated indicators.

5.7 Participation indicators MH-MR teams

The original Robot Participation Rate presents a simple overview of robot involvement: it assumes that all subtasks are of equal duration and complexity, and it does not account for collaborative or shared task execution. To resolve these limitations, we propose a revised metric: the Robot Participation Indicator (RPI), which measures robot contribution based on the actual time spent working of each one, rather than the number of subtasks completed. This individual perspective provides a dedicated value for each robot, expressing how much time that specific robot is active during the process relative to the total time of execution. The new formulation of the Robot Participation Indicator for a single robot is reported below:

$$RPI_{R_i} = \frac{D_{R_i}}{D_{MH-MR}} = \frac{D_{MH-MR} - D_{R_i}^{idle}}{D_{MH-MR}} for\ each\ robot\ i \tag{14}$$

Where:

- D_{R_i} is the time during which the robot i is actively engaged in subtasks.
- D_{MH-MR} is the duration of the MH-MR team assembly process.

The domain is $RPI_{R_i} \in [0,1]$ and it can be interpreted as follows:

- If $RPI_{R_i} = 1$, the robot i works continuously throughout the process (D_{R_i} is equal to D_{MH-MR}).
- If $RPI_{R_i} = 0$, the robot i does not participate at all in the process ($D_{R_i} = 0$).

This version of the indicator supports direct comparisons between robots, helping to identify uneven workload distributions or opportunities for better task balancing.

The proposed RPI is therefore a time-weighted extension of the original Robot Participation Rate, explicitly computed for the analysed MH-MR configurations.

In the first MH-MR configuration of the skateboard assembly case study, both robots are assigned identical tasks and operate in parallel. From the Gantt chart and task allocation (Fig. 6; Table 2), each robot is actively engaged for 93 s during the overall process, which lasts 101 s. Applying Eq. (15), the individual RPI values are:

$$\begin{aligned} RPI_1 &= RPI_2 \\ &= \frac{D_{R_1}}{D_{MH-MR}} = \frac{D_{MH-MR} - D_{R_1}^{idle}}{D_{MH-MR}} \\ &= \frac{101-8}{101} = \frac{93}{101} = 0.92 \end{aligned} \tag{15}$$

These values indicate that each robot is active for 92% of the total task duration. This time-based representation captures the actual effort and temporal engagement of each robot, which would not be visible through simple subtask counting. For example, applying this new metric to the second MH-MR configuration (see Fig. 8), we obtain different RPIs for each robot. They are calculated as follows:

$$\begin{aligned} RPI_1 &= \frac{D_{R_1}}{D_{MH-MR}} = \frac{D_{MH-MR} - D_{R_1}^{idle}}{D_{MH-MR}} = \frac{143-111}{143} = \frac{32}{143} = 0.22 \\ RPI_2 &= \frac{D_{R_2}}{D_{MH-MR}} = \frac{D_{MH-MR} - D_{R_2}^{idle}}{D_{MH-MR}} = \frac{143-0}{143} = \frac{143}{143} = 1 \end{aligned} \tag{16}$$

These additional results highlight how the RPI can identify differences in workload distribution between robots within the same team. In the second configuration, Robot 1 is active for only 22% of the process, whereas Robot 2 is active for 100% of the task. Such a disparity would not be captured by a metric based solely on the number of subtasks completed (see Sect. 4.7), especially if both robots were assigned the same number of tasks. Thus, the RPI provides a more accurate evaluation of robot engagement and identifies opportunities for more effective workload redistribution in MH-MR teams. As with other individual-based metrics, once the RPI for each robot has been calculated, different aggregation strategies can be adopted to derive a single overall indicator for the MH-MR team. These include:

1. Mean RPI, which offers a general overview of robot involvement.
2. Maximum RPI, which highlights the robot with the highest utilisation.

3. Minimum RPI, which helps identify underutilised agents.

As discussed before, the choice of aggregation method should align with the evaluation goals, whether the focus is on overall resource use, efficiency, or identifying performance outliers within the team.

The same formulation can be symmetrically extended to human operators. A Human Participation Indicator (HPI) can be defined for each human agent as:

$$HPI_{H_i} = \frac{D_{H_i}}{D_{MH-MR}} = \frac{D_{MH-MR} - D_{H_i}^{idle}}{D_{MH-MR}} for\ each\ human\ i \tag{17}$$

Where:

- D_{H_i} is the time during which the human *i* is actively engaged in subtasks, obtained from the total process duration and the idle time of each one.
- D_{MH-MR} is the duration of the MH-MR team assembly process.

As before, the domain of HPI_{H_i} is $HPI_{H_i} \in [0,1]$ and it can be interpreted as follows:

- If $HPI_{H_i} = 1$, the human i works continuously throughout the process (D_{H_i} is equal to D_{MH-MR}).
- If $HPI_{H_i} = 0$, the human *i* does not participate at all in the process ($D_{H_i} = 0$).

This allows direct comparison between robotic and human contributions within the same MH-MR configuration. When we apply this new metric to the second MH-MR configuration (see Fig. 8), we obtain a different HPI for each human. They are calculated as follows:

$$\begin{aligned} HPI_1 &= \frac{D_{H_1}}{D_{MH-MR}} = \frac{D_{MH-MR} - D_{H_1}^{idle}}{D_{MH-MR}} = \frac{143-13}{143} = \frac{130}{143} = 0.91 \\ HPI_2 &= \frac{D_{H_2}}{D_{MH-MR}} = \frac{D_{MH-MR} - D_{H_2}^{idle}}{D_{MH-MR}} = \frac{143-111}{143} = \frac{32}{143} = 0.22 \end{aligned} \tag{18}$$

Similar to the RPI, various aggregation strategies can be applied to HPI values, including Mean HPI, Maximum HPI and Minimum HPI. This allows the evaluation to focus on either overall human engagement, peak utilisation, or the identification of underutilised human operators.

In addition to individual participation indicators, a Team Utilisation Indicator (TUI) can be defined as the average active-time ratio across all agents:

$$TUI = \frac{\left(\sum_{i=1}^{n_{rMH-MR}} RPI_i + \sum_{j=1}^{n_{hMH-MR}} HPI_j\right)}{n_{rMH-MR} + n_{hMH-MR}} \tag{19}$$

Where:

- RPI_i and HPI_j are the individual participation rates of robots and humans involved;
- n_{rMH-MR} is the number of robots in the MH-MR team;
- n_{hMH-MR} is the number of humans in the MH-MR team;

The domain of the Team Utilisation Indicator is $TUI \in [0,1]$.

- If $TUI = 1$, all agents are continuously active throughout the process.
- If $TUI = 0$, no agent is engaged in task execution.

The application of TUI to the second MH-MR team configuration (see Fig. 8) involving two humans and two robots is computed as follows:

$$\begin{aligned} TUI &= \frac{RPI_1 + RPI_2 + HPI_1 + HPI_2}{n_{rMH-MR} + n_{hMH-MR}} \\ &= \frac{0.22+1+0.91+0.22}{2+2} = \frac{2.35}{4} = 0.59 \end{aligned} \tag{20}$$

This team-level indicator, derived from individual idle times, provides a synthetic measure of overall resource utilisation, reflecting the average active engagement of team members in MH-MR contexts. In the analysed configuration, the obtained value (TUI=0.59) indicates that, on average, agents are active for around 59% of the total process duration. This reveals a moderate utilisation level and highlights an imbalanced level of participation among team members.

For a complete interpretation, the TUI should be interpreted together with MH-MR Concurrent Activity, that captures full-team synchronisation and parallelism, to avoid potential loss of information when using a single aggregated indicator [18].

5.8 General discussion

To clarify the position of the proposed framework relative to existing literature on the evaluation of MH-MR teams' performance, Table 11 compares recent research studies. Starting from Riedelbauch et al., [50], that we use as a basis to define new KPIs for MH-MR teams, the reported approaches focus on teamwork perspective [38], on team fluency metrics [25] and on multi-robot multi-operator prediction model [8]. This paper proposes reformulated productivity KPIs explicitly extended to MH-MR teams, incorporating team composition, agent-type heterogeneity, task participation, and system robustness considerations within a quantitative framework.

Table 11 Comparison of evaluation dimensions for MH-MR teams in existing frameworks

Analysis dimension	Riedelbauch et al., [50]	Ma et al., [38]	Hoffman, [25]	Boschetti et al., [8]	This document
Productivity times	Productivity metrics (e.g., cooperative speed-up, makespan)	Mission/task duration	Idle time, concurrent activity, functional delay	System-level cycle time and normalized makespan modelling	Productivity KPIs explicitly extended to MH-MR teams
Team size	Primarily dyadic HRT benchmarking	Explicit MH-MR team structures	Dyadic (1 H-1R) model	Multi-robot, multi-operator configurations	Explicitly modelled as a variable within KPI formulation
KPIs on agents	Separate metrics between agents	No explicit agent-differentiated KPIs	No explicit agent-differentiated KPIs	No explicit agent-differentiated KPIs	KPIs with weighted human and robot contributions
Task contribution/Participation	Participation rate (subtask-based) and idle metrics	Not explicitly quantified	Indirectly reflected via idle and concurrency metrics	Indirectly reflected (e.g. collaboration parameter c%) but no explicit participation metric	Time-based agent participation indicator and team utilisation indicator
Reliability/Errors	Robot error rate metrics	Reliability discussed at system/team level	Not explicitly addressed	Not explicitly addressed	Error rate indicators in MH-MR teams
Manufacturing focus	Yes, collaborative assembly systems	No, general HRT domains	No, general HRC domains	Yes, collaborative assembly systems	Yes, collaborative assembly systems
Level of evaluation	Agent level	Teamwork level	Interaction level	System-level	Agent-level and MH-MR team-level
Application field	Not described	Spacecraft maintenance	Turn-taking manipulation task	Prototype collaborative assembly workcell for simulation	Manufacturing assembly

The comparison between frameworks highlights three main aspects. First, while time-based evaluation is common across the considered papers, the level at which it is addressed differs substantially, ranging from interaction-level timing measures to system-level cycle time modelling. Second, explicit modelling of team size and agent-type differentiation within KPI definitions remains very limited. Third, quantitative indicators capturing task participation in MH-MR contexts are not systematically integrated, unlike the proposed formulation.

Regarding the Collaboration Efficiency metric, a simplifying assumption is adopted: agents belonging to the same category (humans and robots) are considered equal in terms of performance and their contribution is assumed to scale linearly with their number. However, in real MH-MR teams, differences in type of robots, skill levels, capabilities, autonomy, or utilisation may lead to potential differentiation within the same agent type. Although the current model defines different weighting factors to capture human and robotic contributions, other weighting strategies could be introduced to consider this variability between each single agent.

Moreover, the proposed KPIs have been developed and validated within manufacturing assembly scenarios, which represent the empirical context of this study. However, the underlying dimensions addressed (e.g. time-based contribution, participation balance, agent-type differentiation, team size) are structural properties of mixed human-robot teams.

Finally, a further examination of the robustness of the proposed KPIs under proportional time variations is provided in Appendix A.

6 Conclusions

Multi-Human Multi-Robot (MH-MR) teams are expected to transform collaborative dynamics across a wide range of industries, but they introduce new levels of complexity and continuous coordination. Compared to dyadic HRC configurations, MH-MR teams usually do not operate on fixed interaction paths: roles, information flows and collaboration dynamics may evolve during task execution and must be managed in real time [44]. In particular, human operators may simultaneously interact with robotic agents and coordinate with other humans, increasing the need for shared situational awareness and effective decision-support tools [41]. These organisational and cognitive challenges underscore the importance of structured, multi-dimensional evaluation strategies. This study makes a contribution to this area by proposing revised productivity metrics for assessing

Table 12 Subtask duration perturbation (±20%) applied to the first configuration of MH-MR team

ID Subtasks	−20% perturbed time (s)	Previous time (s)	+20% perturbed time (s)
S1	23.20	29	34.80
S2	20	25	30
S3	6.40	8	9.60
S4	2.40	3	3.60
S5	4	5	6
S7	38.40	48	57.60
S8	32	40	48

Table 13 Effect of time perturbation on time-based KPIs

Metric	Reference value	−20% perturbed value	Variation (Δ%)	+20% perturbed value	Variation (Δ%)
CE_{MH-MR}	1.15	1.44	+25%	0.96	−17%
H_{RMH-MR}	0.13	0.31	+138%	−0.04	−131%
$D_{H1}^{idle} = D_{H2}^{idle}$	20 s	16 s	−20%	24 s	+20%
$D_{R1}^{idle} = D_{R2}^{idle}$	8 s	6.40 s	−20%	9.60 s	+20%
$MH - MR\ D_{Coop}$	73 s	58.40 s	−20%	87.60 s	+20%
$RPI_1 = RPI_2$	0.92	0.92	-	0.92	-
$HPI_1 = HPI_2$	0.80	0.80	-	0.80	-
TUI	0.86	0.86	-	0.86	-

MH-MR teams in manufacturing that address coordination, task distribution, and type of agents. However, the proposed metrics provide only a partial representation of MH-MR performance. For example, they do not capture directly important dimensions, such as task complexity, human operator's cognitive load, safety, and adaptability. In particular, the metrics do not capture temporal and behavioural dynamics that emerge in MH-MR teams (e.g. operator fatigue, learning curves, inter-agent trust, changes in task allocation, team structure over time). These limitations also reflect the deterministic nature of the present work. The proposed KPIs need to be validated through stochastic task-time modelling. Future research should therefore assess KPIs' robustness under variability and uncertainty, through experiments in controlled and real-world manufacturing environments, as well as simulation-based analyses incorporating stochastic task durations and alternative team configurations.

Further developments may also extend the proposed framework by integrating additional analysis dimensions, like cognitive demand, learning effects, human well-being, and adaptability to disturbances. Ultimately, the integration of these dimensions into real-time adaptive control systems may contribute to dynamically optimising MH-MR team performance.

In summary, this paper establishes a foundation for a more comprehensive, team-oriented assessment approach in human-robot collaboration, helping researchers and practitioners to design MH-MR teams that are fast, productive, well-coordinated, resilient, and human-aware.

Appendix A: Sensitivity analysis under time perturbation

In addition to structural assumptions, robustness under execution variability represents another relevant aspect of KPI validity. As an example, we consider a variation of ±20% in the time of subtasks in the first configuration of MH-MR team described (see Table 12). The human-only baseline remains unchanged (233 s), allowing direct comparison of productivity indicators before and after the variations. The resulting effects on Collaboration Efficiency, Relative Helpfulness, Idle Times, Concurrent activity, Participation indicators and Team Utilisation Indicator are summarised in Table 13.

As shown in Table 13, time-based productivity indicators are directly influenced by the ±20% uniform variation applied to all subtasks. The proportional scaling of task durations leads to corresponding changes in the overall MH-MR completion time, significantly affecting efficiency-oriented metrics such as Collaboration Efficiency and MH-MR Relative Helpfulness. The non-symmetric percentage variation observed in Collaboration Efficiency (e.g. +25% for a 20% reduction in time and −17% for a 20% increase) derives from the reciprocal structure of the time ratio embedded in the metric. As it is inversely proportional to the MH-MR completion time, proportional reductions in duration result in amplified gains, whereas proportional increases generate attenuated losses. Idle times and concurrent activity vary proportionally with task durations, reflecting the uniform temporal perturbation applied to the process. In contrast, participation indicators (RPI and HPI) and the Team Utilisation Indicator (TUI) remain unchanged. As both individual active times and the total process duration are scaled by the same percentage, the ratios that define these indicators remain constant. These results confirm that efficiency metrics are sensitive to global temporal variations, whereas utilisation and participation measures are structurally invariant under proportional time scaling. This distinction reinforces the complementary role of the proposed KPIs in capturing different dimensions of MH-MR team performance.

Author contributions All authors contributed to the study's conception and design. The first draft of the manuscript was written by A. S. Di Matteo and M. Capponi under the supervision of L. Mastrogiacomo and F. Franceschini. All authors read and approved the final manuscript.

Funding Open access funding provided by Politecnico di Torino within the CRUI-CARE Agreement. Not applicable.

Data availability Not applicable.

Declarations

Ethical Approval The authors respect the Ethical Guidelines of the Journal.

Consent to Participate Not applicable.

Consent to Publish Not applicable.

Competing interests The authors declare no competing interests.

References

1. Abrams AMH, Der Pütten AMR (2020) I–C–E framework: concepts for group dynamics research in human-robot interaction: revisiting theory from social psychology on ingroup identification (I), cohesion (C) and entitativity (E). Int J Soc Robot 12(6):1213–1229. https://doi.org/10.1007/s12369-020-00642-z
2. Adel A (2022) Future of industry 5.0 in society: human-centric solutions, challenges and prospective research areas. J Cloud Comput 11(1):40. https://doi.org/10.1186/s13677-022-00314-5
3. Aderinto F, Bhagat Smith J, Giolando M-R, Baskaran P, Adams JA (2024) Improving Human-Robot Team Transparency with Eye-tracking based Situation Awareness Assessment. Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction, 177–181. https://doi.org/10.1145/3610978.3640647
4. Ajoudani A, Zanchettin AM, Ivaldi S, Albu-Schäffer A, Kosuge K, Khatib O (2018) Progress and prospects of the human–robot collaboration. Auton Robot 42(5):957–975. https://doi.org/10.1007/s10514-017-9677-2
5. Angulo C, Chacón A, Ponsa P (2024) Workspace metrics and evaluation. In Cognitive Assistant Supported Human-Robot Collaboration (pp. 77–105). Elsevier. https://doi.org/10.1016/B978-0-44-322135-4.00013-4
6. Bauer A, Wollherr D, Buss M (2008) HUMAN–ROBOT COLLABORATION: a survey. Int J Humanoid Robot 05(01):47–66. https://doi.org/10.1142/S0219843608001303
7. Berx N, Decré W, De Schutter J, Pintelon L (2025) A harmonious synergy between robotic performance and well-being in human-robot collaboration: a vision and key recommendations. Annu Rev Control 59:100984. https://doi.org/10.1016/j.arcontrol.2024.100984
8. Boschetti G, Bottin M, Faccio M, Minto R (2021) Multi-robot multi-operator collaborative assembly systems: a performance evaluation model. J Intell Manuf 32(5):1455–1470. https://doi.org/10.1007/s10845-020-01714-7
9. Bruno G, Antonelli D (2018) Dynamic task classification and assignment for the management of human-robot collaborative teams in workcells. Int J Adv Manuf Technol 98(9–12):2415–2427. https://doi.org/10.1007/s00170-018-2400-4
10. Capponi M, Gervasi R, Mastrogiacomo L, Franceschini F (2024) Assembly complexity and physiological response in human-robot collaboration: insights from a preliminary experimental analysis. Robot Comput Integr Manuf 89:102789. https://doi.org/10.1016/j.rcim.2024.102789
11. Capponi M, Gervasi R, Mastrogiacomo L, Franceschini F (2025) Proposal of a complexity model for human-robot collaboration assembly processes. Robot Comput Integr Manuf 95:103026. https://doi.org/10.1016/j.rcim.2025.103026
12. Correia F, Melo FS, Paiva A (2024) When a robot is your teammate. Top Cogn Sci 16(3):527–553. https://doi.org/10.1111/tops.12634
13. Dahiya A, Aroyo AM, Dautenhahn K, Smith SL (2022) A Survey of Multi-Agent Human-Robot Interaction Systems. arXiv. https://doi.org/10.48550/arXiv.2212.05286
14. De Visser EJ, Peeters MMM, Jung MF, Kohn S, Shaw TH, Pak R, Neerincx MA (2020) Towards a theory of longitudinal trust calibration in human–robot teams. Int J Soc Robot 12(2):459–478. https://doi.org/10.1007/s12369-019-00596-x
15. Edgar G, Aygun A, McWilliams M, Scheutz M (2024) Towards genuine robot teammates. Improving human-robot team performance beyond shared mental models with proactivity
16. Esterwood C, Robert LP (2020) Human Robot Team Design. Proceedings of the 8th International Conference on Human-Agent Interaction 251–253. https://doi.org/10.1145/3406499.3418751
17. European Commission Directorate General for Research and Innovation. (2021). Industry 5.0: Towards a sustainable, human centric and resilient European industry. Publications Office. https://data.europa.eu/doi/https://doi.org/10.2777/308407
18. Franceschini F, Galetto M, Maisano D (2019) Designing Performance Measurement Systems: Theory and Practice of Key Performance Indicators. Springer International Publishing. https://doi.org/10.1007/978-3-030-01192-5
19. Gao F, Cummings ML, Solovey E (2016) Designing for Robust and Effective Teamwork in Human-Agent Teams. https://doi.org/10.1007/978-1-4899-7668-0_9

20. Gervasi R, Mastrogiacomo L, Franceschini F (2020) A conceptual framework to evaluate human-robot collaboration. Int J Adv Manuf Technol 108(3):841–865. https://doi.org/10.1007/s00170-020-05363-1
21. Gombolay MC, Gutierrez RA, Clarke SG, Sturla GF, Shah JA (2015) Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. Auton Robot 39(3):293–312. https://doi.org/10.1007/s10514-015-9457-9
22. Goodrich MA, Jr DRO (2003) Metrics for evaluating human-robot interactions
23. Groom V, Nass C (2007) Can robots be teammates? Benchmarks in human–robot teams. Interact Stud Social Behav Communication Biol Artif Syst 8(3):483–500. https://doi.org/10.1075/is.8.3.10gro
24. Harriott CE, Buford GL, Adams JA, Zhang T (2015) Measuring human workload in a collaborative human-robot team. J Human-Robot Interact 4(2):61. https://doi.org/10.5898/JHRI.4.2.Harriott
25. Hoffman G (2019) Evaluating fluency in human–robot collaboration. IEEE Trans Human-Mach Syst 49(3):209–218. https://doi.org/10.1109/THMS.2019.2904558
26. Hoffman G, Breazeal C (2004), September 20 Collaboration in Human-Robot Teams. AIAA 1st Intelligent Systems Technical Conference. AIAA 1st Intelligent Systems Technical Conference, Chicago, Illinois. https://doi.org/10.2514/6.2004-6434
27. IJtsma M, Ma LM, Pritchett AR, Feigh KM (2019) Computational methodology for the allocation of work and interaction in human-robot teams. J Cogn Eng Decis Mak 13(4):221–241. https://doi.org/10.1177/1555343419869484
28. Iqbal T, Riek LD (2017) Coordination dynamics in multihuman multirobot teams. IEEE Rob Autom Lett 2(3):1712–1717. https://doi.org/10.1109/LRA.2017.2673864
29. Jo W, Wang R, Yang B, Foti D, Rastgaar M, Min B-C (2024) Cognitive Load-based Affective Workload Allocation for Multi-human Multi-robot Teams. arXiv. https://doi.org/10.48550/arXiv.2303.10465
30. Johannsmeier L, Haddadin S (2017) A hierarchical human-robot interaction-planning framework for task allocation in collaborative industrial assembly processes. IEEE Rob Autom Lett 2(1):41–48. https://doi.org/10.1109/lra.2016.2535907
31. Karbouj B, Schuster T, Blumhagen PS, M., Krüger J (2024) Optimizing Human-Robot Collaboration in Industry 5.0: A Comparative Study of Communication Mediums and Their Impact on Worker Well-being and Productivity. 2024 IEEE 6th International Conference on Cognitive Machine Intelligence (CogMI), 182–188. https://doi.org/10.1109/CogMI62246.2024.00032
32. Krüger J, Lien TK, Verl A (2009) Cooperation of human and machines in assembly lines. CIRP Ann 58(2):628–646. https://doi.org/10.1016/j.cirp.2009.09.009
33. Kruijff GJM, Kruijff-Korbayová I, Keshavdas S, Larochelle B, Janíček M, Colas F, Liu M, Pomerleau F, Siegwart R, Neerincx MA, Looije R, Smets NJJM, Mioch T, Van Diggelen J, Pirri F, Gianni M, Ferri F, Menna M, Worst R, Hlaváč V (2014) Designing, developing, and deploying systems to support human–robot teams in disaster response. Adv Robot 28(23):1547–1570. https://doi.org/10.1080/01691864.2014.985335
34. Lamon E, De Franco A, Peternel L, Ajoudani A (2019) A capability-aware role allocation approach to industrial assembly tasks. IEEE Robot Autom Lett 4(4):3378–3385. https://doi.org/10.1109/LRA.2019.2926963
35. Lewis M, Wang H, Chien S-Y, Scerri P, Velagapudi P, Sycara K, Kane B (2010) Teams organization and performance in multi-human/multi-robot teams. 2010 IEEE International Conference on Systems, Man and Cybernetics, 1617–1623. https://doi.org/10.1109/ICSMC.2010.5642379
36. Liau YY, Ryu K (2020) Task allocation in human-robot collaboration (HRC) based on task characteristics and agent capability for mold assembly. Procedia Manuf 51:179–186. https://doi.org/10.1016/j.promfg.2020.10.026
37. Lu Y, Zheng H, Chand S, Xia W, Liu Z, Xu X, Wang L, Qin Z, Bao J (2022) Outlook on human-centric manufacturing towards industry 5.0. J Manuf Syst 62:612–627. https://doi.org/10.1016/j.jmsy.2022.02.001
38. Ma LM, Ijtsma M, Feigh KM, Pritchett AR (2022) Metrics for human-robot team design: a teamwork perspective on evaluation of human-robot teams. ACM Trans Human-Robot Interact 11(3):1–36. https://doi.org/10.1145/3522581
39. Maurtua I, Ibarguren A, Kildal J, Susperregi L, Sierra B (2017) Human–robot collaboration in industrial applications: safety, interaction and trust. Int J Adv Robot Syst 14(4):172988141771601. https://doi.org/10.1177/1729881417716010
40. Merlo E, Lamon E, Fusaro F, Lorenzini M, Carfì A, Mastrogiovanni F, Ajoudani A (2023) An ergonomic role allocation framework for dynamic human–robot collaborative tasks. J Manuf Syst 67:111–121. https://doi.org/10.1016/j.jmsy.2022.12.011
41. Mina T, Kannan SS, Jo W, Min B-C (2020) Adaptive workload allocation for multi-human multi-robot teams for independent and homogeneous tasks. IEEE Access 8:152697–152712. https://doi.org/10.1109/ACCESS.2020.3017659
42. Mordor Intelligence (2026), February 19 *Collaborative Robots Market Size, Share & Forecast, 2031*. Mordor Intelligence. https://www.mordorintelligence.com/industry-reports/collaborative-robot-market
43. Murphy RR, Schreckenghost D (2013) Survey of metrics for human-robot interaction. 2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 197–198. https://doi.org/10.1109/HRI.2013.6483569
44. Natarajan M, Seraj E, Altundas B, Paleja R, Ye S, Chen L, Jensen R, Chang KC, Gombolay M (2023) Human-Robot teaming: grand challenges. Curr Rob Rep 4(3):81–100. https://doi.org/10.1007/s43154-023-00103-1
45. Obi I, Wang R, Jo W, Min B-C (2025) Modeling and Evaluating Trust Dynamics in Multi-Human Multi-Robot Task Allocation. arXiv. https://doi.org/10.48550/arXiv.2409.16009
46. Parks M, Allison M (2023) Autonomy and Fatigue in Human-Robot Teams. 2023 IEEE International Performance, Computing, and Communications Conference (IPCCC), 158–159. https://doi.org/10.1109/IPCCC59175.2023.10253851
47. Petzoldt C, Harms M, Freitag M (2023) Review of task allocation for human-robot collaboration in assembly. Int J Comput Integr Manuf 36(11):1675–1715. https://doi.org/10.1080/0951192x.2023.2204467
48. Ranz F, Hummel V, Sihn W (2017) Capability-based task allocation in human-robot collaboration. Procedia Manuf 9:182–189. https://doi.org/10.1016/j.promfg.2017.04.011
49. Reinerman-Jones LE, Hughes N, D'Agostino A, Matthews G (2019) Human performance metrics for the nuclear domain: a tool for evaluating measures of workload, situation awareness and teamwork. Int J Ind Ergon 69:217–227. https://doi.org/10.1016/j.ergon.2018.12.001
50. Riedelbauch D, Höllerich N, Henrich D (2023) Benchmarking teamwork of humans and cobots—an overview of metrics, strategies, and tasks. IEEE Access 11:43648–43674. https://doi.org/10.1109/ACCESS.2023.3271602
51. Shukla A (2024) Performance Metrics for Collaborative Robots: A Literature Review. https://doi.org/10.31224/3964
52. Simões AC, Pinto A, Santos J, Pinheiro S, Romero D (2022) Designing human-robot collaboration (HRC) workspaces in industrial settings: a systematic literature review. J Manuf Syst 62:28–43. https://doi.org/10.1016/j.jmsy.2021.11.007
53. Singer S, Akin D (2011), July 17 A Survey of Quantitative Team Performance Metrics For Human-Robot Collaboration. 41st

International Conference on Environmental Systems. 41st International Conference on Environmental Systems, Portland, Oregon. https://doi.org/10.2514/6.2011-5248

54. Steinfeld A, Fong T, Kaber D, Lewis M, Scholtz J, Schultz A, Goodrich M (2006) Common metrics for human-robot interaction. Proc 1st ACM SIGCHI/SIGART Conf Human-Robot Interact 33–40. https://doi.org/10.1145/1121241.1121249
55. Tabrez A (2022) Descriptive and prescriptive visual guidance to improve shared situational awareness in human-robot teaming
56. Villani V, Pini F, Leali F, Secchi C (2018) Survey on human–robot collaboration in industrial settings: safety, intuitive interfaces and applications. Mechatronics 55:248–266. https://doi.org/10.1016/j.mechatronics.2018.02.009
57. Wang R, Zhao D, Gupte A, Min B-C (2023) Initial Task Assignment in Multi-Human Multi-Robot Teams: An Attention-enhanced Hierarchical Reinforcement Learning Approach. arXiv. https://doi.org/10.48550/arXiv.2310.04979
58. Yuan Z, Wang R, Kim T, Zhao D, Obi I, Min B-C (2025) Adaptive Task Allocation in Multi-Human Multi-Robot Teams under Team Heterogeneity and Dynamic Information Uncertainty. arXiv. https://doi.org/10.48550/arXiv.2409.13824
59. Zhang S, Chen Y, Zhang J, Jia Y (2020) Real-Time Adaptive Assembly Scheduling in Human-Multi-Robot Collaboration According to Human Capability. 2020 IEEE Int Conf Rob Autom (ICRA) 3860:3866. https://doi.org/10.1109/ICRA40945.2020.9196618

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.