

A Multifaceted User Study for the Teaching-Learning-Prediction-Collaboration Framework in Human-Robot Collaborative Tasks

Omar Obidat, Garrett Modery, Weitian Wang*, *Senior Member, IEEE*, Xiwang Guo, *Senior Member, IEEE*, and Mengchu Zhou, *Fellow, IEEE*

Abstract —As the implementation of robotics systems in modern industries becomes more commonplace, the desire to streamline and simplify humans’ interaction with them is highly needed. Human-robot collaboration frameworks have made strides towards the goal to facilitate shared tasks in human-robot teams. Such methods as Learning from Demonstration (LfD) show great potential in well performing collaborative tasks. To boost LfD’s capacity, our previous study has developed a novel Teaching-Learning-Prediction-Collaboration (TLPC) framework for robots to learn from human demonstrations, customize their task strategies according to humans’ personalized working preferences, predict human intentions, and assist humans in collaborative tasks. In this work, we conduct a multifaceted user study to evaluate it in real-world human-robot collaborative tasks. Participants of this user study are from diverse age groups with varying educational backgrounds and genders. Seven assessment metrics are developed to comprehensively evaluate the performance of TLPC through *t*-tests. A controlled human-robot collaborative experiment without TLPC is also conducted. This study seeks to observe and analyze the subjective feelings and feedback of the participants using TLPC when they perform collaborative tasks with a robot via periodic surveys given throughout the experiment. Our research outcomes help us gather insights into and create catalysts for the construction and optimization of human-robot interactive systems in advanced manufacturing contexts. They can be leveraged to improve human-robot collaboration quality, manufacturing productivity, human safety, and ergonomics.

I. INTRODUCTION

Manufacturing industries have experienced a substantial increase in the utilization of robotics in recent years as markets continue to grow [1, 2]. Human-robot collaboration techniques can effectively help manufacturing sectors remain flexible and productive to anticipate and handle frequent market changes [3]. Both parties function as a team in which each member makes the most efficient use of their abilities. The robot’s strength, precision, and repeatability are combined with the human’s ability to make plans, decisions, and reactions, such that the goal is accomplished faster and safer. The partnership has been shown to be both productive as well as cost-effective in studies [4, 5].

Recent studies have sought to observe the effectiveness of a concept that is relatively new to the fields of robotics and automation, specifically in the area of robot learning from human demonstrations [6, 7]. Implementing such collaborative methods is a process to ensure a robot’s efficacy, human capacity, and collaborative performance. Elements of consideration include ease of use, convenience, safety, and overall human comfort with the machine. These are further described and embodied by factors such as the speed

of a robot, for instance, which has a direct impact on humans’ perception of safety [8]. It is important to consider that a user who is dissatisfied and uncomfortable with a teammate – whether a human or robot – would see that overall productivity suffers as a result. Motivated by these needs, our work [9] proposes a Teaching-Learning-Prediction-Collaboration (TLPC) framework for the robot to learn from human demonstrations to cooperatively accomplish a task in line with each human’s personalized working preference. Through it, the robot can learn different strategies to complete a task, understand ongoing human actions, predict human intentions, and automatically adjust its motions to collaborate with its human partner in the shared tasks.

Building on the developed TLPC framework, we conducted a multifaceted user study in this work to evaluate the performance of TLPC in real-world human-robot collaborative tasks. Participants engaged with the robot for a collaborative assembly task with multiple parts using or not using TLPC. After each human-robot collaborative experiment, a survey is completed where users answer questions related to the sociability of the robot, as well as the trust, comfort, and fluency felt during the experiment. Using this data, we can draw conclusions related to the effectiveness of human-robot collaboration and humans’ overall satisfaction with the process, which can then be used to further tailor this form of collaboration and the developed TLPC framework to improve the efficiency of human-robot teams.

Users’ perception and opinion of their robot teammate are crucial to the success of human-robot collaboration. Studies such as those mentioned provide desired insights into different methods of teaching, demonstrations, processes, and their strengths. Additionally, some of these efforts [10, 11] have evaluated their participants’ subjective responses to their robot partners. What remains to be comprehensively analyzed, however, are the numerous factors that have direct effects on a human’s collaborative partnership with a robot. These factors include the sociability of a robot, its movement speed, the fluency of the collaboration, etc. While these particular elements may have been observed in other studies, they are often isolated or generalized. Our use of multifaceted assessment metrics (task efficiency, collaboration safety, coding efforts reduction, collaboration fluency, sociability, robot response speed, and overall comfort) to be examined together through *t*-tests allows a comprehensive evaluation of TLPC and its impact on the collaboration process between a human and robot. Our study ultimately seeks to present insights into human-robot collaborative processes and catalyze the optimization of human-robot interactive system implementation in advanced manufacturing contexts.

II. RECAP OF THE TLPC FRAMEWORK

Effective collaboration between two parties involves having an understanding of each other’s intentions. A fruitful partnership is built on the foundational idea of

O. Obidat, G. Modery, and W. Wang are with the School of Computing, Montclair State University, Montclair, NJ 07043 USA. (corresponding author: wangw@montclair.edu)
M. Zhou and X. Guo are with the Department of Electrical and Computer Engineering, New Jersey Institute of Technology Newark, NJ 07102, USA.

achieving a common goal, and knowledge of how each member seeks to achieve that goal is required. This mutual understanding offers a greater degree of fluency in executing the task at hand and accomplishing the desired outcome when compared to alternative scenarios in which this understanding is not present. In human-robot interaction, the user should know what to expect from their robot teammate. Likewise, permitting the robot to predict the user's intentions has immense potential in improving the speed, accuracy, and general efficiency of the process. In this context, we have proposed a TLPC framework that demonstrates how a robot may learn from human demonstrations to perform a given task [9]. As it operates, the robot becomes more accurate and customizable with its user's working preferences. The approach uses a finite-state machine model that permits the robot to learn alternative pathways to a common goal. This, combined with the long short-term memory (LSTM) algorithm [12], enables the robot to act dynamically as it predicts the user's next actions.

In TLPC, the teaching segment consists of three stages in which the robot observes the actions taken by the user and utilizes an LSTM network to process the data in order to locate the objects in it. The second segment involves the stages related to the robot's ability to learn the task. From the previous stage, the operation information of the objects is determined and used in an algorithm to create a finite-state machine with the purpose of building the robot's task knowledge. The states within this finite-state machine represent the various positions of the objects and their configurations, while the transitions indicate the movements of the objects from one state to another state. Lastly, the final stage of TLPC is the prediction-collaboration segment. In this process, the robot predicts the next state the user will transition to by using the state machine developed from the learned experience and assists the user in reaching it by executing the next most suitable action (e.g., passing the needed part to the user). This design provides the robot with some autonomy to "think" by itself to help its human partner accomplish the task well.

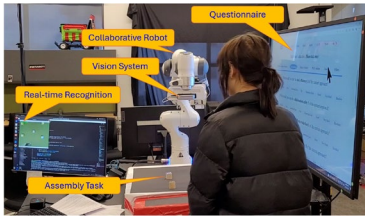


Fig. 1. Experimental setup for the user study.

III. USER STUDY DESIGN FOR TLPC FRAMEWORK

A. Experimental Platform

In this work, we create a high-precision experimental setup to evaluate our proposed model with human participants, as shown in Fig. 1. The setup includes a collaborative robot, five parts with a letter on each part spelling the word "ROBOT" acting as a collaborative assembly product, an intel RealSense depth camera used as the robot's vision system, a TV screen for participants to answer our questionnaire, a workspace designed for collaboration, and two workstations; one to execute the TLPC model and another to collect participants' answers. The experiment involves tasking a human participant with assembling the word "robot" in different sequences to

teach the robot how to complete the assembly process in future human-robot collaborative tasks. The robot we are using is a Franka-Emika Panda. The data processing and model execution are handled by a ThinkStation P520 workstation. We used the Robot Operating System (ROS) for effective management of the robot system and human-robot collaboration process [9].

B. Task Design

For our task, we ask participants to assemble the word "robot" by using five wooden parts in different assembly sequences. Despite being a basic task, we believe that this serves as a proof of concept, as a five-part assembly task has 120 potential patterns to choose from. In addition, the verification of this task can be scaled to more complex tasks. The collaboration can be split into two separate subtasks. The first subtask implements TLPC in which a participant teaches the robot two different assembly patterns that he/she prefers, and the robot learns the assembly strategies from human demonstrations. Participants then test if the robot can predict their assembly preference and collaborate properly with them. The second subtask that we implement is a controlled experiment involving a hard-coded approach in which a participant does not teach and the robot cannot learn, but has the most common patterns pre-programmed into it. This allows the robot to collaborate if it recognizes the pattern that participants create; and if it fails to recognize it, the robot stops operating.

TABLE I. ASSESSMENT METRICS FOR THE TLPC MODEL

Metric	Description
Task Efficiency	How efficiently the robot performs the task.
Collaboration Safety	The safety level during the collaboration.
Coding Efforts Reduction	The degree to which the approach reduces the need for programming.
Collaboration Fluency	How smoothly the robot interacts with the participant.
Collaboration Sociability	The perceived social aspect of interacting with the robot.
Robot Response Speed	The speed at which the robot responds to the participant's actions.
Overall Comfort	The general comfort level of the participant while interacting with the robot.

The primary goal of the task design is to test TLPC against current hard-coded approaches and allow participants to answer questions that may demonstrate if TLPC provides any improvement to the process of human-robot collaboration. To assess the efficacy of the TLPC and hard-coded approach, participants are asked to complete a detailed survey post-interaction for each subtask. This survey consists of fourteen questions, each with a nine-point Likert scale ranging from 'Extremely Poor' to 'Excellent'. The first seven questions are for the first sub-task (with TLPC), and the remaining seven are for the second sub-task (with the hard-coded approach). The assessment metrics are described in Table I. Participants answer these questions for both approaches, allowing for a comparative evaluation and analysis of TLPC.

C. Data Collection

With the nine-point Likert scale, participants have a wide range of options to rate these evaluation metrics because we hope to capture exactly how they feel and quantify the smallest changes for more accurate results. The data is collected by using an HTML webpage that is developed through Flask, a microweb framework. The data is saved in a local database

that has two tables named Users and QuestResponse. The collected data is split into three sections. The first one includes the participants' demographic information. The second one contains the answers to the seven questions regarding the first sub-task. The third section includes the answers to the last seven questions regarding the second sub-task. We have successfully collected data from 86 participants including 36 male participants (41.86%), 48 female participants (55.81%), and 2 other participants (2.33%).

IV. DATA ANALYSIS APPROACHES

To analyze the data collected from the participants, we have conducted a multifaceted and comparative evaluation through three forms of the t -test approach [13, 14], which is a statistical tool crucial for understanding differences in group behaviors and outcomes in user studies. The three forms used are Student's t -test, Welch's t -test, and paired t -test. Welch's t -test does not require the standard deviations of the two groups being compared to be the same, whereas Student's t -test assumes that they are, which may not necessarily be the case in practice. This allows it to be a more robust analysis method when comparing data. We utilize both Welch's and Student's t -tests to examine the differences between them and observe the paired t -test as well. Paired t -test is particularly useful in scenarios such as ours in which one subject is measured at two different points in time, or under two different conditions.

To compute the Student's t -test, three main data values are essential, including Mean Difference (the difference between the means of both datasets), Standard Deviations (the standard deviations of each dataset), and Number of Values (the number of data samples in each dataset). In our study, when analyzing samples from distinct subtasks, we calculate this t -test value as follows:

$$t = (m_A - m_B) / \sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}, \quad (1)$$

where m_A and m_B are the means of groups A and B, separately, n_A and n_B denote their sizes, respectively, and S^2 is an estimator of the pooled variance of the two groups:

$$S^2 = \frac{\sum_{i=1}^{n_A} (x_i - m_A)^2 + \sum_{j=1}^{n_B} (x_j - m_B)^2}{n_A + n_B - 2}. \quad (2)$$

The degrees of freedom (d) is $d = n_A + n_B - 2$, where x_i and x_j represent the i_{th} and j_{th} sample values of groups A and B, respectively. While applying the t -test, we adhere to four critical assumptions including Measurement Scale (the data should be on an ordinal/continuous scale), Sample Representativeness (the sample must accurately represent the groups from which it is drawn), Normal Distribution (the data should exhibit a normal, bell-shaped distribution curve), and Homogeneous Variance (this assumption holds if the standard deviations of both datasets are approximately equal, indicating similar variability) [24, 25].

The second t -test that we employ in our analysis is Welch's t -test. This test is preferred over Student's t -test because it permits the standard deviations of the subsets to be different from each other [26]. For our data, this is particularly useful considering our standard deviations vary

between the two concerned approaches. Welch's t -test is expressed as:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s_1^2 + s_2^2}}, \quad (3)$$

where $\bar{x}_1$ and $\bar{x}_2$ represent the sample means of groups 1 and 2, and s_1 and s_2 are the differing standard deviations of groups 1 and 2, respectively. The degrees of freedom of Welch's t -test, v , is defined as:

$$v \approx \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{s_1^4}{N_1^2 v_1} + \frac{s_2^4}{N_2^2 v_2}}, \quad (4)$$

where N_1 and N_2 denote the data sizes of groups 1 and 2.

The paired t -test is used when data exists in matched pairs. In our data, we may use this approach because subjects evaluated both approaches during the experiment, permitting us to compare their subjective thoughts on them. We calculate its t statistic as follows:

$$t = \frac{\bar{x}_D}{s_D / \sqrt{n}}, \quad (5)$$

where $\bar{x}_D$ represents the average difference between paired values, and s_D indicates the standard deviation of the differences. This test is slightly simpler than the previous two tests, and its degrees of freedom is calculated as $n - 1$, where n is the number of matched pairs.

Gathering t -values permits the determination of p -values, which indicates the probability of observing results as extreme as or more extreme than the ones seen. Small values suggest that there is evidence against the null hypothesis, which implies that there is no significant difference between the groups being compared. Our hypotheses are defined as Null Hypothesis (no significant difference between the two groups being compared) and Alternative Hypothesis (there does exist a significant difference between the two groups).

It should be noted that p -values can be determined by using t -values and t table [15]. This table permits an estimation of the p -value by lookup with the calculated t -value and v -value. However, more specific calculations must be done to establish a more precise p -value. These calculations often make use of the Cumulative Distribution Function, which describes the probability that a random variable X will take a value less than or equal to a value x . This has direct applications in our t -tests and subsequent calculation of p -values [16]. To establish these hypotheses within a given context, a statistical significance level of 0.05 (5%) has been chosen. In other words, p -values less than 0.05 place the data in favor of rejection of the null hypothesis.

V. RESULTS AND ANALYSIS

A. Human-Robot Collaborative Process Analysis

The results of our real-world human-robot collaboration are related to the categories detailed in Table I. The process by which one of the female participants develops her subjective positions on the interactions that she takes part in is detailed in Fig. 2. In Fig. 2(a), she teaches the robot to execute the collaborative task in her preferred assembly order as the robot observes her task strategy using the TLPC

model through its vision system. This assembly strategy is learned, and when the learning process is completed as shown in Fig. 2(b), the robot is prepared to work and assist her. Having learned her preferred assembly order, the robot proceeds to help its human teammate complete the task collaboratively by grabbing and handing the predicted parts to her. In Figs. 2(c-d), this prediction takes place as the robot examines the current state of the collaborative task using the task knowledge that it has built during the learning phase. Once a decision is made regarding which part should be operated given the current context in addition to the constructed task knowledge, the robot moves to grab and hand that part to her. Figs. 2(e-f) show the collaboration process as it occurs, with the participant taking a part, allowing the robot to take the next and hand it to her, and continuing the process until the task is completed.

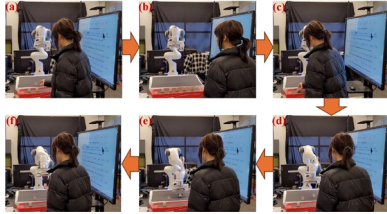


Fig. 2. Human-robot collaboration using the TLPC framework.

To contrast the human-robot collaboration efforts of task completion utilizing TLPC and that without it, Fig. 3 presents how it differs from the process detailed in Fig. 2. In Figs. 3(a-c), the robot acts similarly to how it operates during the collaborative approach using TLPC. The difference, however, is that it does not observe and learn the user's preferred assembly sequence, but rather has been pre-programmed to function with predefined orders. It observes the first part taken and attempts to determine the order that the user is attempting to assemble the parts based upon that input and which predefined sequence it could potentially be. Once it makes a decision, it begins the collaborative process by providing the next part based upon its assumption. This works well when the user's sequence coincides with one of the predetermined ones. However, that is not always the case. As mentioned before, with five parts to choose from in any order, there exist one hundred and twenty possible combinations. In a scenario where the robot fails to recognize the sequence in which the participant begins to take the parts in, it ceases the collaborative process. This instance is shown in Figs. 3(d-f) where the robot is unable to collaboratively complete the task with the user, ending the interaction.

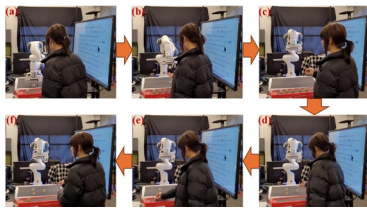


Fig. 3. Human-robot collaboration without using the TLPC model.

B. Evaluation of Task Efficiency

Three types of t -tests are performed on the data to gather insight into the relationship between the subsets collected. Fig. 4 presents a comparative analysis of participants' perceived efficiency of the task during each of the

interactions using TLPC and the hard-coded approach. Using three types of t -tests – Welch's, Student's, and paired – we gain the t -values, v -values, and p -values shown in Table II. It should be noted as well that a mean difference of 1.6 exists between the data of the two different approaches. From these results, it is reasonably concluded that there exists a statistically significant difference between the responses for these data from the two methods. This is due to all p -values being far lower than the chosen significance level threshold of 0.05. When observing Fig. 4, this difference becomes apparent, as it shows that an overwhelming majority of participants have selected "excellent" and "very good" for the approach utilizing TLPC, whereas the non-TLPC approach has received primarily "good" responses and lower ones.

TABLE II. T-VALUES, V-VALUES, AND P-VALUES OF T-TESTS ON THE EVALUATION OF TASK EFFICIENCY

t -tests	t	v	P
Student's t -test	6.76	163	2.29e-10
Welch's t -test	6.71	138.85	4.4e-10
Paired t -test	7.72	80	2.82e-11

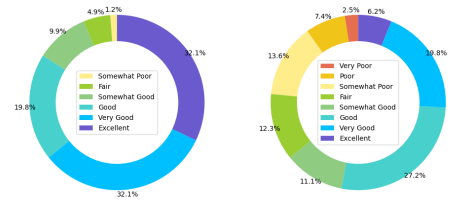


Fig. 4. Participants' ratings for the task efficiency question with (left) and without (right) the TLPC model.

TABLE III. T-VALUES, V-VALUES, AND P-VALUES OF T-TESTS ON THE EVALUATION OF COLLABORATION SAFETY

t -tests	t	v	P
Student's t -test	3.38	170	9.13e-4
Welch's t -test	3.38	140	9.5e-4
Paired t -test	4.27	85	4.98e-5

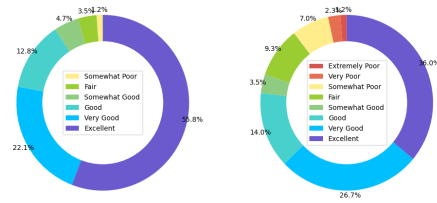


Fig. 5. Participants' ratings for the collaboration safety question with (left) and without (right) the TLPC model.

C. Evaluation of Collaboration Safety

Fig. 5 provides a visual representation of the responses to the metric regarding perceived collaboration safety throughout each interaction. Performing Student's, Welch's, and paired t -tests produces the t -, v -, and p -values in Table III. Regardless of which test is, the values are indicative of a statistically significant difference between the two approaches when comparing them to the chosen significance level of 0.05. The mean difference between the approaches is about 0.81, which is slightly lower than that for task efficiency. This, along with comparatively higher p -values, is due to a somewhat closer relationship between the two groups in this data than those in the previous section. While the majority of responses for both sections fell into the "excellent" category, the approach using TLPC contained,

on average, much higher ratings. In other words, participants seem to feel that safety was higher when the robot has watched and learned how to perform the task, rather than having its instructions hardcoded in.

D. Evaluation of Coding Efforts Reduction

Fig. 6 demonstrates the responses for a perceived reduction in coding efforts for a given approach. All three t -tests, including Welch's, Student's, and paired, produce a p -value far below our designated significance level value of 0.05. The t - and v -values are also obtained, as shown in Table IV. Thus, these values serve as strong evidence against the null hypothesis, which states that there is not a statistically significant difference between the two human-robot collaboration approaches. In other words, it is highly unlikely that the evaluation results would have taken this form if the null hypothesis were true. Evidently, most participants selected "excellent" for the TLPC approach (left), but only "very good" or lower for the non-TLPC approach (right). With means of ~ 6.83 and ~ 5.36 , a mean difference of 1.47 is produced, also showing a notable difference between the two approaches. From this data, we can conclude that participants generally feel that permitting the robot to learn and collaborate offers a significant reduction in coding efforts.

TABLE IV. T-VALUES, V-VALUES, AND P-VALUES OF T-TESTS ON THE EVALUATION OF CODING EFFORTS REDUCTION

t -tests	t	v	P
Student's t -test	5.99	170	1.18e-8
Welch's t -test	5.99	144.09	1.56e-8
Paired t -test	7.72	85	2.06e-11

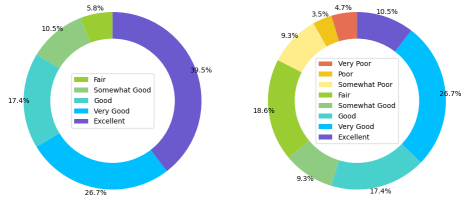


Fig. 6. Participants' ratings for the coding efforts reduction question with (left) and without (right) the TLPC model.

TABLE V. T-VALUES, V-VALUES, AND P-VALUES OF T-TESTS ON THE EVALUATION OF COLLABORATION FLUENCY

t -tests	t	v	P
Student's t -test	6.43	167	1.28e-9
Welch's t -test	6.42	147.14	1.79e-9
Paired t -test	7.72	83	2.33e-11

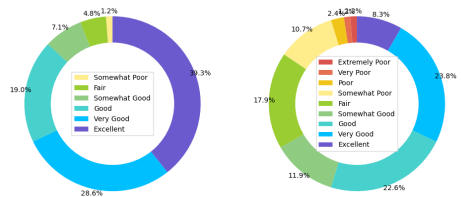


Fig. 7. Participants' ratings for the collaboration fluency question with (left) and without (right) the TLPC model.

E. Evaluation of Collaboration Fluency

Fig. 7 compares the responses to the fluency of human-robot collaboration for each approach. It is evident from the resulting data that participants feel that the approach utilizing TLPC provides a more fluent experience than the

one without, with far more "excellent" responses. The t -test results for this metric are very much the same as those for previously discussed survey questions. Our three methods in Table V, Welch's, Student's, and paired t -tests, produce p -values much lower than the 5% significance level value. With these values, it is highly likely that, on average, there exists a statistically significant difference between participant's perspectives from one approach to the other. Additionally, a mean difference of 1.49 exists between them, suggesting that many participants feel that TLPC offers a more fluent experience.

F. Evaluation of Collaboration Sociability

Fig. 8 presents the responses to the perceived sociability of the robot throughout the human-robot collaboration experience. As presented in Table VI, Welch's, Student's, and paired t -tests are utilized to examine the relation between the responses. From the evaluation results, it is clear that the observed p -values are indicative that the two dataset's differences are likely statistically significant when held against the 5% significance level. Additionally, there exists a mean difference of about 1.2, allowing us to reasonably conclude that participants feel that the approach using TLPC provides the robot with a more social demeanor when compared to the non-TLPC approach.

TABLE VI. T-VALUES, V-VALUES, AND P-VALUES OF T-TESTS ON THE EVALUATION OF COLLABORATION SOCIABILITY

t -tests	t	v	P
Student's t -test	4.88	170	2.37e-6
Welch's t -test	4.89	153.61	2.57e-6
Paired t -test	6.43	85	6.99e-9

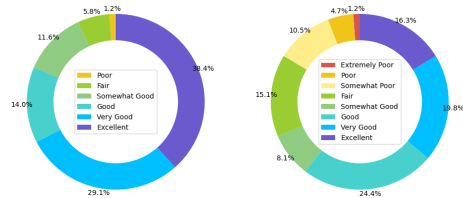


Fig. 8. Participants' ratings for the collaboration sociability question with (left) and without (right) the TLPC model.

TABLE VII. T-VALUES, V-VALUES, AND P-VALUES OF T-TESTS ON THE EVALUATION OF ROBOT RESPONSE SPEED

t -tests	t	v	P
Student's t -test	3.46	170	6.7e-4
Welch's t -test	3.46	168.12	6.7e-4
Paired t -test	4.76	85	7.7e-6

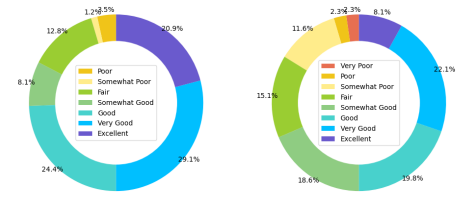


Fig. 9. Participants' ratings for the robot response speed question with (left) and without (right) the TLPC model.

G. Evaluation of Robot Response Speed

Fig. 9 shows the responses from participants for each approach regarding the response speed of the robot during the human-robot collaboration. The t -, v -, and p -values of the three t -tests for this metric are presented in Table VII.

Despite being low and suggesting a statistical difference between the two approaches, the comparatively higher p -values indicate that participants feel that there is less of a difference between the approaches regarding response speed than they do for some of the other metrics, such as fluency. The mean difference for this data is approximately 0.86, suggesting that, on average, subjects typically drop less than a point on the scale when rating the non-TLPC approach. These results, however, indicate that the participants prefer the TLPC framework.

H. Evaluation of Overall Comfort

Fig. 10, shows the comparison of the participants' overall comfort during each of the collaborative interactions. It is easy to immediately see a large difference between the approaches from the results, but this difference becomes more apparent when examining the t -, v -, and p -values, as shown in Table VIII. These p -values, similar to those seen in other metrics, exist far below the designated significance level of 0.05. This provides a strong indication that there exists a statistically significant difference between the two methods used in this experiment in the context of overall comfort during human-robot interaction. Notably, the mean difference between these data is 0.94, suggesting that, on average, subjects rate each interaction nearly one point differently than the other. We can conclude that subjects typically feel more comfortable interacting with the robot when it learns and collaborates with them dynamically, rather than simply being hardcoded to perform its designated task.

TABLE VIII. T-VALUES, V-VALUES, AND P-VALUES OF T-TESTS ON THE EVALUATION OF OVERALL COMFORT

t -tests	t	v	P
Student's t -test	4.23	169	3.76e-5
Welch's t -test	4.22	141.65	4.29e-5
Paired t -test	5.5	84	3.93e-7

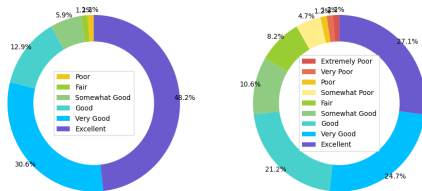


Fig. 10. Participants' ratings for the overall comfort question with (left) and without (right) the TLPC model.

VI. CONCLUSIONS AND FUTURE WORK

We have conducted a multifaceted user study to evaluate the TLPC framework in real-world human-robot collaborative tasks. Participants of this user study are diverse in their age, educational background, and gender. We have designed seven assessment metrics to comprehensively evaluate the performance of the TLPC framework. Through a controlled experiment, we have compared the performance of human-robot collaboration with and without it. The evaluation results suggest the competitive strengths of the developed TLPC framework to ensure fruitful human-robot collaboration. In addition, we hope that the findings from this user study can serve as catalysts for the construction and optimization of human-robot interactive systems to improve the productivity of human-robot partnerships in advanced manufacturing and remanufacturing [17, 18]. Our next work is to gather more

data and perform their deep analysis in various aspects to gain new findings to help advance TLPC.

ACKNOWLEDGMENT

This work is supported in part by the National Science Foundation under Grants CMMI-2138351 and CNS-2117308.

REFERENCES

- [1] M. Javaid, A. Haleem, R. P. Singh, and R. Suman, "Substantial capabilities of robotics in enhancing industry 4.0 implementation," *Cognitive Robotics*, vol. 1, pp. 58-75, 2021.
- [2] A. Buerkle *et al.*, "Towards industrial robots as a service (IRaaS): Flexibility, usability, safety and business models," *Robotics and Computer-Integrated Manufacturing*, vol. 81, p. 102484, 2023.
- [3] C. Zheng, X. Qin, B. Eynard, J. Bai, J. Li, and Y. Zhang, "SME-oriented flexible design approach for robotic manufacturing systems," *Journal of Manufacturing Systems*, vol. 53, pp. 62-74, 2019.
- [4] W. Wang, Y. Chen, R. Li, Z. Zhang, V. Krovi, and Y. Jia, "Human-Robot Collaboration for Advanced Manufacturing by Learning from Multi-Modal Human Demonstrations," *Recent Advances in Industrial Robotics*, World Scientific Publishing, pp. 87-116, 2020.
- [5] W. Wang, R. Li, Z. M. Diekel, Y. Chen, Z. Zhang, and Y. Jia, "Controlling Object Hand-Over in Human-Robot Collaboration Via Natural Wearable Sensing," *IEEE Transactions on Human-Machine Systems*, vol. 49, no. 1, pp. 59-71, 2019.
- [6] W. Wang, R. Li, Y. Chen, Z. M. Diekel, and Y. Jia, "Facilitating Human-Robot Collaborative Tasks by Teaching-Learning-Collaboration From Human Demonstrations," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 2, pp. 640-653, 2018.
- [7] W. Wang, R. Li, Y. Chen, Y. Sun, and Y. Jia, "Predicting Human Intentions in Human-Robot Hand-Over Tasks Through Multimodal Learning," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 2339-2353, 2022.
- [8] S. Haddadin, A. Albu-Schäffer, and G. Hirzinger, "Safety Evaluation of Physical Human-Robot Interaction via Crash-Testing," in *Robotics: Science and Systems*, 2007, vol. 3, pp. 217-224.
- [9] O. Obidat, *et al.*, "Development of a Teaching-Learning-Prediction-Collaboration Model for Human-Robot Collaborative Tasks," in *2023 IEEE International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2023: IEEE, pp. 728-733.
- [10] A. Apraiz, G. Lasa, and M. Mazmela, "Evaluation of user experience in human-robot interaction: a systematic literature review," *International Journal of Social Robotics*, vol. 15, no. 2, pp. 187-210, 2023.
- [11] R. A. Søraa, G. Tøndel, M. W. Kharas, and J. A. Serrano, "What do older adults want from social robots? A qualitative research approach to human-robot interaction (HRI) studies," *International Journal of Social Robotics*, vol. 15, no. 3, pp. 411-424, 2023.
- [12] J. Bi, X. Zhang, H. Yuan, J. Zhang, and M. Zhou, "A hybrid prediction method for realistic network traffic with temporal convolutional network and LSTM," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 3, pp. 1869-1879, 2021.
- [13] D. W. Zimmerman and B. D. Zumbo, "Rank transformations and the power of the Student t test and Welch t -test for non-normal populations with unequal variances," *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, vol. 47, no. 3, p. 523, 1993.
- [14] E. Hedberg and S. Ayers, "The power of a paired t -test with a covariate," *Soc Sci Res*, vol. 50, pp. 277-291, 2015.
- [15] R. H. Browne, "The t -test p value and its relationship to the effect size and $P(X > Y)$," *Am Stat*, vol. 64, no. 1, pp. 30-33, 2010.
- [16] W. T. Shaw, "Sampling Student's T distribution-use of the inverse cumulative distribution function," *Journal of Computational Finance*, vol. 9, no. 4, p. 37, 2006.
- [17] S. Lou, *et al.*, "Human-Cyber-Physical System for Industry 5.0: A Review From a Human-Centric Perspective," *IEEE Transactions on Automation Science and Engineering*, pp. 1-18, 2024.
- [18] X. Guo, *et al.*, "Balancing Human-robot Collaborative Disassembly Line by Using Dingo Optimization Algorithm," in *Proc. IEEE International Conference on Human-Machine Systems*, 2024, pp. 1-6.