Design of a Virtual Task to Understand the Nature of Collaboration Between Autistic and Neurotypical Adults in Workplace Using Multimodal Data

Mahrukh Tauseef[1](✉), Ashwaq Zaini Amat[1], Deeksha Adiani[1], Spencer Hunt[1], Makayla Honaker[2], Amy Swanson[2], Amy Weitlauf[2], and Nilanjan Sarkar[1]

[1] Electrical and Computer Engineering Department, School of Engineering, Vanderbilt University, Nashville, TN, USA
mahrukh.tauseef@vanderbilt.edu

[2] Treatment and Research Institute for Autism Spectrum Disorder, Vanderbilt University Medical Center, Nashville, TN, USA

Abstract. Secure employment is an essential milestone of adulthood due to its association with self-sufficiency, personal achievement, and financial stability. Over the years, autistic population has been greatly affected by high unemployment rates, which is often attributed to their social skills deficit. Existing vocational facilities employ long-term intervention programs that rely on qualitative data to assess collaboration and social skills. However, qualitative data might be prone to error and bias. In addition, long-term qualitative analysis requires the presence of an observer throughout the program, which can result in high operational cost of such training facilities. This paper proposes the design of a collaborative virtual environment that includes LEGO building tasks that are mapped to nine dimensions of collaboration, which can be assessed by collecting and analyzing multimodal data captured during the tasks. A feasibility study with 6 participants was conducted to test the system. The results showed that the system was able to correctly capture multimodal data that corresponded to different dimensions of collaboration. Self-reported surveys reflected that the proposed system fostered a collaborative environment for the participants to exercise their social skills.

Keywords: Collaboration assessment · Social skills training · Employment for autistic adults · Autism · Multimodal data · VR-based HCI Systems · Serious games

1 Introduction

Securing employment has become one of the most essential milestones of adulthood due to the self-sufficiency, financial independence, and social acceptance that it has to offer [1]. It is often considered a measure of self-worth, achievement, and societal status [2]. Thus, unemployment can have drastic effects on the

M. Antona and C. Stephanidis (Eds.): HCII 2022, LNCS 13309, pp. 410–426, 2022.
https://doi.org/10.1007/978-3-031-05039-8_30

mental health, development of interpersonal relationships, and financial stability of an individual [3]. In addition, it can increase financial strain on the families of unemployed individuals as well as the government that has to come up with resources to provide support [2].

Autistic individuals are amongst one of the most vulnerable populations who suffer from high unemployment rates. Autism spectrum disorder (ASD) is a complex developmental condition with persistent deficit in social interactions, restrictive interests, and/or repetitive behavior [4]. According to a survey conducted in 2017, about 5.5 million adults had ASD [5]. Despite their cognitive abilities and technical skills, the high unemployment rate amongst autistic adults has been a major concern over the years. A study conducted in 2015 suggested that at least 42% of autistic individuals in their early 20's were unemployed. These findings have been supported by many studies since then [6–8]. In addition, the ever-increasing prevalence of autism amongst children (1 in 44 reported in 2021 [9]) suggests that there will be even more autistic adults seeking employment in the coming years.

Social skills are becoming increasingly important in the current labor market. According to a report led by Microsoft Corporation, the capability to communicate and collaborate with colleagues is among the core skills that they are looking for in future employees [10]. Unfortunately, the high unemployment rate amongst autistic population is often attributed to the challenges many of them face with social interaction and collaboration [2]. Not only does it make it difficult for them to secure employment, but it also becomes challenging for them to maintain it. Wei et al. [6] observed that the average employed autistic adult seems to maintain employment for half the time as compared to an average neurotypical adult. Thus, several government and private vocational intervention programs have been set up to provide autistic individuals necessary support in order to secure employment [2].

Ke et al. [11] conducted a survey and classified existing social skills intervention techniques available for autistic adults and youth into three categories: direct instruction interventions, naturalistic interventions, and technology-based interventions. Direct instruction intervention depends on the presence of a coach or a facilitator who could instruct, role-play or provide feedback [11]. Naturalistic interventions consist of activities designed for autistic individuals and their peers that puts them in a social situation that targets certain skills and provide corrective feedback based on their actions [11]. However, such activities require a well-structured curriculum and constant presence of a facilitator over a series of sessions. Many governmental vocational programs are often underfunded and understaffed to afford a working model whereas private vocational programs that may be able to afford such training can be expensive [2].

Even though direct and naturalistic intervention techniques are more common, technology-based intervention techniques are becoming more popular over the years [12]. This technique supports implementation of conventional methods through video modeling or Virtual Reality (VR) based serious games [11]. They allow the implementation of a variety of easily personalized social scenarios in a controlled environment and reduce added stress by eliminating social scenarios that might induce high anxiety [13]. Since autistic individuals commonly have an affinity for computers, they are more likely to be motivated during technology-based intervention [12,13].

Currently, most intervention techniques, whether they are conventional or technological, rely on surveys and questionnaires filled by facilitators and/or the participants for the assessment of skill training [11]. Such qualitative analysis is prone to bias. In addition, it relies on the constant surveillance of an observer, which can add to the staff's workload as well as the operational cost of a program. Quantitative assessment of collaboration can not only provide unbiased feedback, but it can also reduce the cognitive and financial pressure on these programs. Advancement of human computer interface (HCI) design allows for the development of a VR application that can collect multimodal data that can be used to assess collaboration quantitatively. However, such a system requires mapping of different dimensions of collaboration with measurable data from an HCI system that can be a reliable indicator of collaboration.

In this paper, we propose the design of a naturalistic collaborative virtual environment (CVE) shared by two users that (1) includes two specialized LEGO [14] building tasks that are mapped to nine different collaboration dimensions, and (2) utilizes multimodal data - speech, eye gaze, task progression, and controller input - to determine if the proposed system can capture collaborative interaction between participants.

This paper is structured as follows. Section 2 presents an overview of existing VR-based HCI systems that have been developed for assessing collaboration and social skills. This is followed by the design of our proposed CVE in Sect. 3. Section 4 presents the experimental setup for the feasibility study. The results of the study are shown in Sect. 5 followed by a discussion in Sect. 6 and concluding remarks in Sect. 7.

2 Related Work

Numerous VR-based interventions designed for autistic individuals for social skills training have reported significant improvement in social interaction. Didehbani et al. [15] designed a study to investigate the impact of VR-based social skills training on children with ASD. They used a virtual world software called Second Life™ where participants went through different social settings as virtual avatars. Participants interacted with an avatar of a clinician who played different social roles in these scenarios. The results reported significant improvement in the social skills of the children after five weeks of intervention. Another study conducted in Hong Kong reported improvements in social interaction specifically in social reciprocity when a group of children went through real life based social scenarios using a Cave Automatic Virtual Environment [16]. The children were able to interact with one another as well as virtual objects in the system.

Additionally, several VR-based serious games have been designed that specifically focus on collaborative skills training. Silva et al. [17] designed a game for a multitouch table in which young ASD participants worked in pairs to put the uniform on a player together. The game was designed to offer training of four different collaboration patterns: active sharing, passive sharing, joint performance, and unrestricted interaction. The system was evaluated by therapists who observed the participants as they went through different phases of the game. The results showed that the system encouraged the participants to collaborate by showcasing the interactive situations and intentions that corresponded to the four expected patterns of collaboration. Sara et al. [18] designed an intelligent agent that would perform different goal oriented as well as cooperative turn-taking activities with an autistic child. These activities were mapped to different dimensions of collaboration such as reciprocal interaction, sharing intentions, sharing emotions, and other nonverbal cues such as eye gaze. The results of the training reported improvement in collaborative skills.

However, all the aforementioned studies were conducted with autistic children. There are few social and collaborative skills training programs that were designed specifically for autistic adults and even fewer designed for collaborative skill training in a workplace. Kandalaft et al. [12] conducted a version of a study [15] mentioned above with autistic adults and reported improved social skills. One of the most notable VR-based systems for collaborative training focusing on workplace was designed by Microsoft. Their neurodiversity hiring program includes an adaptation of Minecraft that allows candidates to go through different social scenarios that requires team building [19]. However, like all the earlier works mentioned above, the assessment of collaboration relied on the observation of a personnel.

In addition to observations, specialized tests as well as surveys were commonly used to assess and evaluate collaboration. One study utilized multimodal data to assess collaboration skills of autistic population. Alozie et al. [20] presented the design of a real-world collaborative task mapped to different dimensions of collaboration whose assessment relied on the analysis of multimodal data such as verbal communication, head pose, gestures, and eye gaze. However, the collaborative task was designed for autistic children and was mapped to a limited number of dimensions of collaboration.

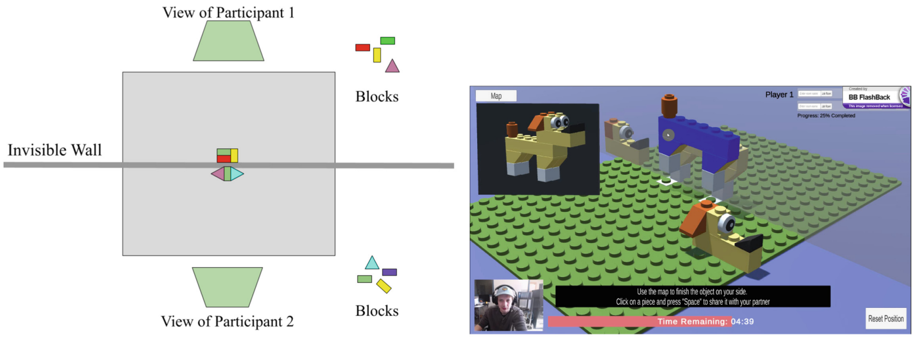
Motivated by the existing work in the design of VR-based intervention systems that have shown potential for improvement in collaborative skills, we present a novel VR-based collaborative system that addresses some of the current challenges. Our proposed system incorporates a series of LEGO building tasks in a virtual environment to foster natural collaboration between two adult participants over a network. This increases the portability of the system by allowing participants to access it from different locations. The building tasks are mapped to nine different dimensions of collaboration that are important in a workplace. The multimodal data collected during the collaborative game can be used to draw inference about the degree of collaboration between the two participants based on the strategic design of the tasks that relies on collaboration for completion.

3 Design

The design of the CVE-based assessment tasks can be divided into three subsections. Section 3.1 goes through the design of the block-building tasks. Section 3.2 provides an overview of the architecture of the collaborative virtual environment, and Sect. 3.3 presents different dimensions of collaboration and their mapping to the tasks.

3.1 Design of Collaborative Tasks

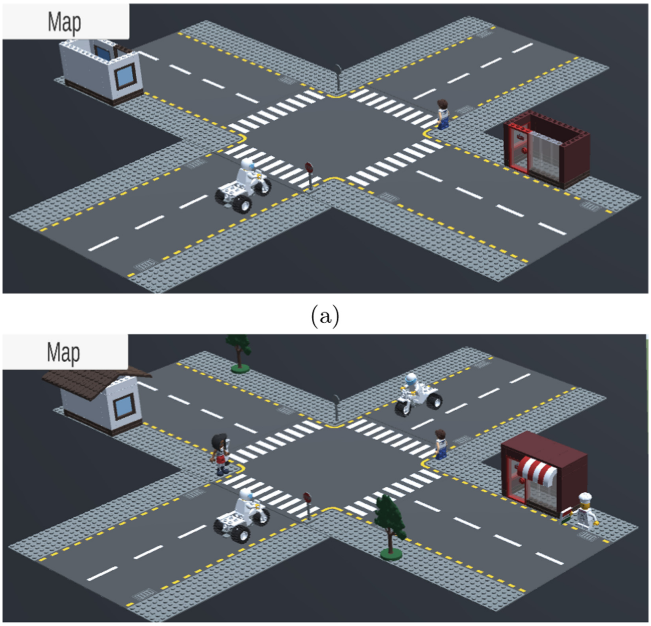
Unity [21], a commercial game development software, was used to design LEGO-inspired block-building tasks in an environment where two participants sit across from one another with a LEGO mat between them (see Fig. 1). An invisible impenetrable wall was used to divide the LEGO mat such that each participant had access to one-half of the mat. All the LEGO pieces were randomly distributed amongst both participants. They were expected to finish their halves of the structure within a time limit by using the pieces on their side and requesting their partner to share required pieces that they do not have access to.



(a) (b)

Fig. 1. A schematic representation of the top view of the building task layout (b) A snapshot of one of the levels of the game with two participants interacting with one another. The interactive map (top left corner), video of player 2 (bottom left corner), and timer bar (bottom center) can be clearly seen.

The task included two tutorials and two levels. The goal of Tutorial A was to acquaint both players with the controls and features of the task independently whereas Tutorial B was designed for them to practice building a simple object in a shared environment. Both the tutorials included elaborate written instructions for each step. Level 1 required the players to build a piano together whereas Level 2 required them to build a scene from a road intersection together by using an interactive map. The map was provided as an aid to both participants that displayed the object or scene they were building. The map allowed them to go through different layers of an object by pressing the *Page up* and *Page down* keys. (see Fig. 2).



(a)

(b)

Fig. 2. (a) Second layer of the interactive map for level 2 (b) Final layer of the interactive map for level 2

The complexity of the tasks (i.e., number of pieces, shape of pieces, type of task, etc.) was increased from Level 1 to Level 2. The variation in the levels was introduced to encompass all the dimensions of collaboration mentioned in Sect. 3.3. Section 3.2 shows the implementation of a collaboration environment that allows the task to run synchronously over a network for both players while collecting multimodal data.

3.2 System Architecture

The proposed CVE system was implemented to support peer-to-peer communication between two people. Some of the necessary components needed for each participant's setup includes a personal computer with a mouse and keyboard to interact with the task, a webcam and a headset that allows the two players to interact with each other, and an eye gaze tracker that can keep track of where the person is looking. Tobii EyeX Eye trackers [22] were used to collect the eye gaze data and Logitech [23] webcams were used to stream the video of each participant.

Figure 3 illustrates the three main components of the CVE-based system. A Network Communication Layer (NCL) links the participants to a shared virtual environment. This network layer was established using the WebRTC API [24]. WebRTC is an open-source real-time communication API that allows for cross platform multimedia communication between nodes [25]. The built-in AV transmission channel was used to exchange audio-visual data between the two players. WebRTC's data channel was used to update and synchronize participants' interactions with each other and the environment. This includes movements of the block pieces made by either of the participants.

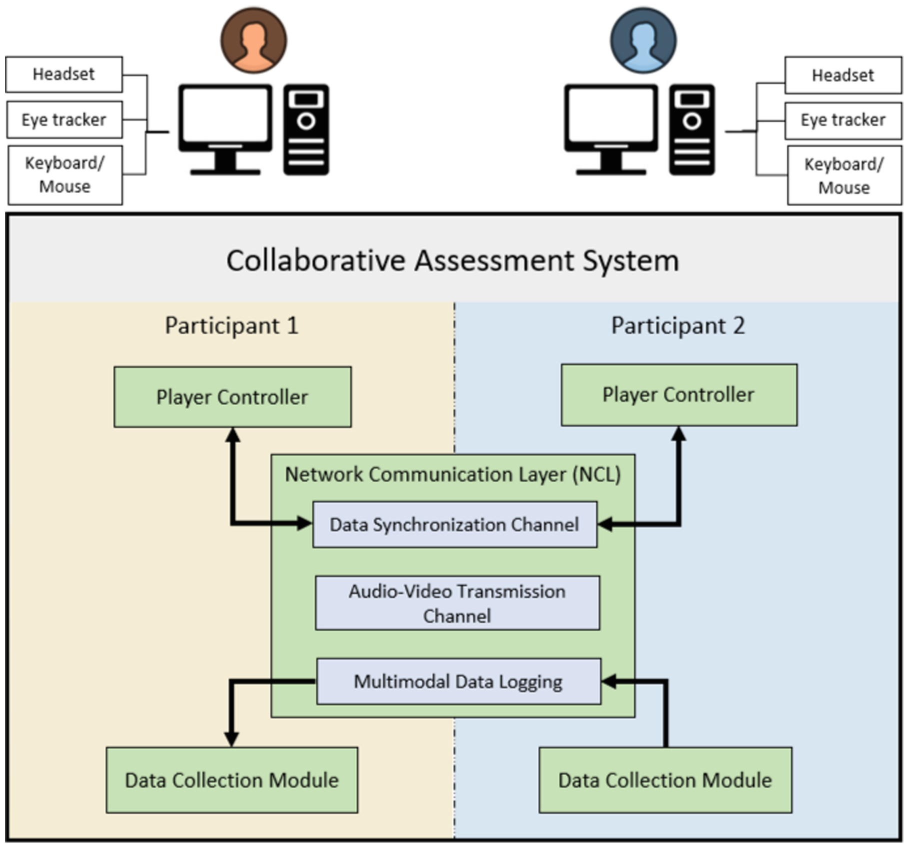


Fig. 3. Collaborative assessment system architecture

Outside of the NCL, there is a Player Controller component and a Data Collection Module for each participant. The Player Controller holds the logic for keyboard buttons presses and mouse clicks when participants manipulate the block pieces. The Data Collection Module stores multimodal input data from participants during the interactions. This includes transcribed speech, gaze points, region of interests (ROIs), keyboard button presses, and time to complete the task. Microsoft Azure's Speech-to-text service [26] was used to transcribe the speech of each participant. The multimodal data collected on each player's side was stored every second with timestamps in a log file on Player 1's PC via WebRTC's data channel. Table 1 defines each entry of the log file.

Table 1. Complete list of multimodal data recorded in the log file from both players

Multimodal data	Description
Timestamp	Instance of time the data was logged
Label	Player 1 or Player 2
Text	Transcribed Speech
Utterance	Number of words uttered
Duration	Duration of the sentence that was uttered
X_Gaze_Point	X-coordinate of where the participant was looking on the screen
Y_Gaze_Point	Y-coordinate of where the participant was looking on the screen
Focused_Object	Name of the Object the participant was looking at
Total_Score	Percentage of pieces latched by both participants
Individual_Score	Percentage of pieces latched by each participant
Piece_Latched	Name of the piece that was just latched
Piece_Shared	Name of the piece that was just shared
Shared_Count	Number of pieces shared by each player
Brick_Selected	Name of the piece that a participant is currently interacting with
Brick_Select_Duration	Amount of time a participant interacted with a certain piece
Active_Effort_Bool	A Boolean that indicates active interaction with the system
Map_Interact_Bool	A Boolean that indicates if a participant is interacting with the map
Time_Remaining	Amount of time remaining once the game ends

3.3 System Measures and Collaboration Matrix Scheme

Although technology-based collaborative intervention has been explored, limited studies looked at the features or dimensions of collaborations that can reflect important communication skills that are essential for workplace collaboration. After extensive research, Meier et al.'s [27] rating scheme for computer-supported collaboration was adapted to design a collaboration matrix that can be mapped to the multimodal data. This matrix was then used to design the levels such that all the collaboration dimensions can be measured by analyzing relevant multimodal data (see Table 2). Designing the levels this way allows for setting varying expectations for collaboration at each level such that participants can get focused on the task at hand without feeling overwhelmed. Setting such expectations also serves as a marker while analyzing the multimodal data of the participant to see if all the expected dimensions of collaboration were exercised.

Table 2. Nine different dimensions of collaboration along with multimodal data that is relevant to each dimension and an example from task design that maps to the dimension

Collaboration dimension	Data relevant to the dimension	Example from the task design that maps to the dimension
Sustaining Mutual Understanding: Do the participants understand each other?	Speech	When an explanation is offered for a question asked by their partner (Tutorial B, Level 1, and Level 2)
Dialogue Management: How are the participants following the etiquette of a conversation?	Speech, eye gaze	When a question is asked and the participant is responding. When the participant is looking at their partner while talking (Tutorial B, Level 1, and 2)
Information Pooling: How often are the participants referring to and/or sharing available information	Speech, Controller Input, Eye gaze	When the participant is talking about or following the instructions. When the participant is looking at or interacting with the map (all levels)
Reaching Consensus How are they coming to a decision together?	Speech	When the participants are discussing if a piece needs to be shared (Tutorial B, Level 1 & 2)
Time Management: How concerned are they about the time constraint?	Speech, Eye Gaze	When the participant looks at the timer and/or mentions if the time is running out (all levels)
Technical Coordination: How are the participants handling technical dependencies?	Speech, Controller Input, Task Progression	When the participants are sharing pieces with each other (Tutorial B, Level 1, and Level 2)
Reciprocal Interaction: Comparable input towards the completion of the task	Controller Input, Task Progression	If the participant is matching the pace and progression of the task on their easily (Tutorial B, Level 1, and Level 2)
Task Division: How was the task divided?	Speech	If the participants came up with a scheme to breakdown, divide and complete their tasks (Tutorial B, Level 1, and Level 2)
Individual Task Orientation: An individual's commitment to complete the task	Speech, Eye Gaze, Task Progression, and Controller Input	How well is a participant doing in finishing the task on their end (all levels)

4 Experimental Setup

A feasibility study was set up to test the designed system. Three autistic adults were paired with 3 typically developed (TD) adults to work on the task together. An ASD-TD pairing was chosen to match real-world circumstances. It is more likely that an autistic individual will end up in a company with a significantly higher number of typically developed individuals. All six participants identified as male within the age range of 18–25 years (mean age = 22.0, SD (Standard Deviation) = 2.16).

The system was set up on two personal computers placed in two different rooms in a lab setting. Both participants were seated in front of their PC's after they signed and handed in a voluntary participation consent form. A screen recording of the participants' display was captured along with the aforementioned multi-modal data. The participants were informed about all the data that was being collected. The study was approved by the Institutional Review Board (IRB) of the authors' university.

Figure 4 shows snapshots of all the levels of the task. Each participant went through Tutorial A to get acquainted with the building tasks independently. Ten minutes were allotted for this level, which was more than the time needed to complete it so that the participants can focus on learning the system without worrying about the time constraint. Support was offered if the participants faced any technical difficulty in this stage. Following Tutorial A, the participants went through Tutorial B, Level 1, and Level 2 in a collaborative environment. Five, 8, and 10 min were allotted for each level, respectively. The task took about 30 min to complete and was followed by a qualitative survey to be filled by each participant.

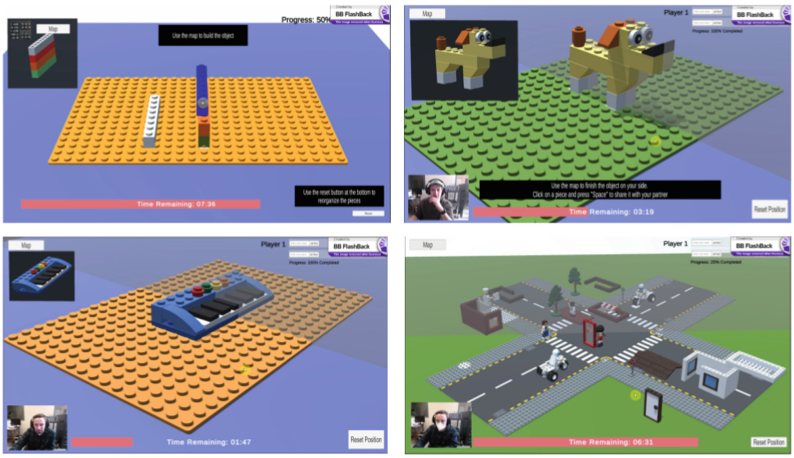


Fig. 4. Snapshot of Tutorial A (top left), Tutorial B (top right), Level 1 (bottom left), and Level 2 (bottom right) from the study

5 Results

All three pairs were able to successfully complete the entire task within the time limits. The log files for all the levels were successfully retrieved. An offline analysis of multimodal data was performed to show that the system can capture different dimensions of collaboration. Some snapshots of the logged data are shown below to demonstrate such mapping.

Sustaining Mutual Understanding: Transcribed speech showed different instances of conversation that presented mutual understanding. Keywords and key phrases like *yes*, *yeah*, *no*, *nope*, *I agree/do not agree* were used to look for such instances. Figure 5 shows a snapshot of a dialogue exchange between two participants that maps to sustaining mutual understanding.

Timestamp	Label	Text
11:36:34	p2	it looks like all that's left is putting the two keys down.
11:36:36	p1	yeah.

Fig. 5. A snapshot of a dialogue exchange between two participants that displays mutual understanding

Dialogue Management. Dialogue management was tracked by using the transcribed speech as well as the utterance. Bidirectional communication was successfully tracked in the log file i.e., whenever a participant started a conversation, it was always reciprocated by their partner. Figure 6 shows a snapshot of a conversation between a pair of participants when they entered the collaborative environment.

TimeStamp	Label	Text	Utterance	Duration
12:23:43	p2	hey how you doing?	4	00:00.1
12:23:44	p1	hey good.	2	00:01.7
12:23:44	p2	good, let's see if we can figure this one out.	10	00:00.1

Fig. 6. A snapshot of a dialogue exchange between two participants that showcases dialogue management

Utterance recorded the number of words said by each participant in a sentence. Table 3 below shows the mean and standard deviation of the number of words uttered per sentence for two pairs. A t-test was performed to see if there was a significant difference in the number of words uttered per sentence during the task by each participant in a pair. For pair 1, no significant difference was reported ($p > 0.05$) whereas a strong statistically significant difference was reported ($p < 0.0001$) for words uttered by each participant in pair 3. The ASD participant uttered fewer words per sentence as compared to the TD participant. Data from pair 2 was not available for analysis as the utterance was not entirely captured. However, this issue was fixed before the system was tested with pair 3.

Table 3. Mean and standard deviation (STD) of words uttered per sentence by each participant in a pair during the task. P-value of a t-test performed to see if there is a significant difference between the number of words uttered

Pair	Pair 1		Pair 3	
	ASD	TD	ASD	TD
Mean	7.05	5.15	3.99	6.16
STD	5.98	4.11	3.28	5.44
p-value	0.076		<0.0001	

Information Pooling: Exchange and sharing of information were represented by the transcribed speech, gaze data, and the Boolean value set for any interaction with the map using the controller input. Figure 7 shows a snapshot of an instance of information pooling from the speech. Figure 10 shows the *Map_Interact_bool* which is set to 1 whenever a participant interacts with the map.

timestamp	player	utterance
11:35:55	p1	you can go through different layers of the map using page up and page down.
11:35:57	p2	oh, really.
11:35:59	p1	oh yes you can.
11:36:00	p1	yeah.
11:36:00	p2	page up page down.
11:36:02	p1	page up or page down.
11:36:06	p2	oh

Fig. 7. Snapshot of transcribed conversation between two participants in a pair that shows an instance of information pooling

Reaching Consensus: Only a few verbal instances of consensus were noticed among the participants. Keywords including *yes*, *you're/you are right*, and *I agree* were used to look for any signs of consensus. Figure 8 shows a snapshot of a conversation that reflects agreement.

Time Management: None of the data recorded display any concerns regarding the time limit. No reference to time management was found in any of the conversations nor was any eye gaze data recorded on the time bar.

Technical Coordination. Since the task completion is dependent on participants' ability to share relevant pieces with each other, the expected technical

TimeStamp	Label	Text
12:41:01	p1	i believe you have another cyclist behind this house that's yours.
12:41:02	p2	yep, you're right. let me knock down this house real quick.

Fig. 8. Snapshot of transcribed conversation between two participants in a pair that shows an instance of reaching consensus

coordination was recorded by the system. Keywords and key phrases like *can you send*, *share/pass* or *I need* were used to look for instances in the transcribed speech where a participant is asking their partner to share a relevant piece. Figure 9 shows a snapshot of a conversation between two participants. One participant is asking for a piece which is followed by their partner's verbal affirmation and action of sharing the piece that was recorded as *Piece_Shared* by using controller input.

TimeStamp	Label	Text	Total_Score	Individual_	Piece_Latched	Piece_Shared
12:41:26	p1	i'll need a tree.	65	60		
12:41:27	p2	there you go, ok and this thank you.	65	70		
12:41:27	p1		65	60		
12:41:28	p2		65	70		P010

Fig. 9. A snapshot of multimodal data recorded in the log file that shows an instance of technical coordination

Another measure of technical coordination was time remaining which was logged as soon as a pair was done with the level. Coordination can be assessed by measuring how long did it take a pair to go through different tasks.

Reciprocal Interaction: Reciprocal interaction was measured by comparing the individual score of each participant, number of pieces shared, and record of any active participation on their side. Figure 10 shows snapshot of a log file where *Individual_Score*, *Shared_Count*, and *active_Effort_Bool* can be used as indicators of reciprocal interaction.

TimeStamp	Label	Text	Total_ Score	Individual _Score	Shared_ Count	active_Effort_ Bool	Map_Interact_ bool
12:41:36	p1		65	60	5	0	0
12:41:37	p2	then yeah, where's the other cyclist you see? have you seen?	65	70	4	1	0
12:41:37	p1		65	60	5	0	0
12:41:39	p2		65	70	4	0	1
12:41:39	p1		65	60	5	0	0
12:41:40	p2	do you see it?	65	70	4	0	0
12:41:40	p1	do you see it behind the white house? you'll see his helmet.	65	60	5	1	0
12:41:41	p2		65	70	4	0	0
12:41:43	p1		65	60	5	1	0
12:41:44	p2	oh yeah. oh i can't grab him yeah, keep him over.	65	70	4	0	0

Fig. 10. A snapshot of multimodal data recorded in the log file shows an instance of reciprocal interaction

Task Division: No instances of task division were observed through the data analysis.

Individual Task Orientation: Individual task orientation can be easily extra cted from the multimodal data. Figure 11 shows all the multimodal data collected from controller input, eye gaze detection, and task progression that can be used to track individual task orientation using the labels *p1* and *p2* for each participant.

TimeStamp	Label	Duration	X_Gaze_ Point	Y_Gaze_ Point	Focused_ Object	Individual_ Score	Shared_ Count	Brick_ Select	Brick_Select_ Duration	active_Effort_ Bool	Map_Interact_ bool
12:41:34	p2	00:00.6	1883.21	105.158	Fixed3	70	4			1	0
12:41:35	p2	00:01.0	1743.97	424.477		70	4			0	0
12:41:36	p2	00:00.7	1743.97	424.477	P012	70	4			0	0
12:41:37	p2	00:00.7	344.291	580.056		70	4	P012	00:02.7	1	0
12:41:38	p2	00:02.4	1106.23	526.198		70	4	P012	00:02.7	1	0
12:41:39	p2	00:02.4	1106.23	526.198		70	4			0	1
12:41:40	p2	00:02.4	1106.23	526.198		70	4			0	0

Fig. 11. A snapshot of multimodal data recorded for 10 s that can be used to assess individual task orientation for participant 2 based on Label = p2

The survey results obtained from each participant can be found in Table 4.

Table 4. Survey results obtained from all participants

Survey question	Scale	Mean	STD
Experience playing the game	1 = Not enjoyable at all, 10 = Very enjoyable	6.7	2.79
Difficulty of the game	1 = Very difficult, 10 = Very easy	6.8	1.62
Communication in the beginning	1 = Not comfortable, 10 = Very comfortable	8.3	1.54
Communication towards the end	1 = Not comfortable, 10 = Very comfortable	9.4	0.85
Allotted time	1 = Not enough, 10 = Too much	6.3	2.10

6 Discussion

The objective of this study was to design a collaborative virtual task and to show that the designed system is capable of recording multimodal data that maps to nine dimensions of collaboration that are essential for collaboration in the workplace. As presented in the previous section, the multimodal data captured by the system can successfully represent mutual understanding, dialogue management, information pooling, consensus, technical coordination, reciprocal interaction, and individual task orientation. However, the system was unable to record any instances of time management or task division.

The amount of time allotted might be one of the reasons why participants were not concerned about time management. This assumption is supported by the results of the survey where participants rated more than enough time when asked for the allotted time. In addition, the system was unable to capture any task division because the existing levels were designed with the block pieces divided between the two players at the beginning of the task shown on the interactive map. Design of an additional *freestyle* level without an interactive map will give the participants more room for decision making and task division as they build something together.

Survey data showed that the participants found the system fairly enjoyable with moderate levels of difficulty (see Table 4). The participant also reported that they were more comfortable communicating with their partners towards the end of the task as compared to the beginning of the task. However, the number of participants is not enough to attribute the increased level of comfort to the design of the system.

A notable observation made from the data analysis was that it could be used to perceive participant's conditions or states. For example, with the multimodal data, we can observe whether a participant was in idle state (i.e., there was no active contribution), struggling (i.e., they were actively participating but making no progress), or in steady state (i.e., they were making progress) over a fixed period of time. This information can be used to design and train an intelligent agent that can offer support catering to the different states (idle or struggling) and upgrade the assessment system to allow for collaborative training.

Only three pairs of participants were enrolled in the study because the objective was to present a proof of concept. Future work includes the design of an intelligent agent that could analyze multimodal data to detect the dimensions of collaboration. The detected dimensions can be compared with the expected collaborative activity for every level to rate collaboration.

7 Conclusion

The motivation behind this work was the unwavering high unemployment rate of autistic population that is often attributed to their social skills deficit. Existing vocational centers are unable to provide cost-effective long-term collaborative skills training because of the added workload on their staff to facilitate and assess collaboration. Existing VR-based systems designed for social skills training mainly focus on autistic children and still depend on personnel for evaluation of collaboration. Our objective was to design a VR-based collaborative virtual task that could assess different dimensions of collaboration using multimodal data. The task was designed to foster a natural collaborative space for an autistic and neurotypical adult to replicate a social interaction that many autistic adults end up having in a workplace. Results of the feasibility study showed that multimodal data can successfully capture different dimensions of collaboration mapped to the task. Future work includes the development of an intelligent agent that can be trained on this multimodal data to automatically detect different dimensions of collaboration.

Acknowledgement. We are grateful for the support provided by NSF grants 1936970 and 2033413 for this research. We would also like to thank the Vanderbilt Treatment and Research Institute for Autism Spectrum Disorders (TRIAD) team; Aislynn Kiser, Katie Wiggins Gregory, and Kelly Luo for their expert advice on interventions for autistic individuals. The authors are solely responsible for the contents and opinions expressed in this manuscript.

References

1. Chen, J.L., Leader, G., Sung, C., Leahy, M.: Trends in employment for individuals with autism spectrum disorder: a review of the research literature. Rev. J. Autism Dev. Disord. **2**(2), 115–127 (2015)
2. Solomon, C.: Autism and employment: implications for employers and adults with ASD. J. Autism Dev. Disord. **50**(11), 4209–4217 (2020)
3. Paul, K.I., Moser, K.: Unemployment impairs mental health: meta-analyses. J. Vocat. Behav. **74**(3), 264–282 (2009)
4. Autism spectrum disorder. https://www.nimh.nih.gov/health/topics/autism-spectrum-disorders-asd
5. Dietz, P.M., Rose, C.E., McArthur, D., Maenner, M.: National and state estimates of adults with autism spectrum disorder. J. Autism Dev. Disord. **50**(12), 4258–4266 (2020)
6. Wei, X., et al.: Job searching, job duration, and job loss among young adults with autism spectrum disorder. J. Vocat. Rehabil. **48**(1), 1–10 (2018)
7. Roux, A.M., Rast, J.E., Shattuck, P.T.: State-level variation in vocational rehabilitation service use and related outcomes among transition-age youth on the autism spectrum. J. Autism Dev. Disord. **50**(7), 2449–2461 (2020)
8. Seagraves, K.: Effective job supports to improve employment outcomes for individuals with autism spectrum disorder. J. Appl. Rehabil. Couns. (2021)
9. Executive summary, December 2021. https://www.cdc.gov/ncbddd/autism/addm-community-report/executive-summary.html
10. 21st century skills and the workplace. https://www.gyli.org/wp-content/uploads/2014/02/21st_century_skills_Gallup.pdf
11. Ke, F., Whalon, K., Yun, J.: Social skill interventions for youth and adults with autism spectrum disorder: a systematic review. Rev. Educ. Res. **88**(1), 3–42 (2018)
12. Kandalaft, M.R., Didehbani, N., Krawczyk, D.C., Allen, T.T., Chapman, S.B.: Virtual reality social cognition training for young adults with high-functioning autism. J. Autism Dev. Disord. **43**(1), 34–44 (2013)
13. Pandey, V., Vaughn, L.: The potential of virtual reality in social skills training for autism: bridging the gap between research and adoption of virtual reality in occupational therapy practice. Open J. Occup. Ther. **9**(3), 1–12 (2021)
14. Official lego®. https://www.lego.com/en-us
15. Didehbani, N., Allen, T., Kandalaft, M., Krawczyk, D., Chapman, S.: Virtual reality social cognition training for children with high functioning autism. Comput. Hum. Behav. **62**, 703–711 (2016)
16. Yuan, S.N.V., Ip, H.H.S.: Using virtual reality to train emotional and social skills in children with autism spectrum disorder. London J. Primary Care **10**(4), 110–112 (2018)
17. Silva, G.F.M., Raposo, A., Suplino, M.: PAR: a collaborative game for multitouch tabletop to support social interaction of users with autism. Proc. Comput. Sci. **27**, 84–93 (2014)
18. Bernardini, S., Porayska-Pomsta, K., Smith, T.J.: Echoes: an intelligent serious game for fostering social communication in children with autism. Inf. Sci. **264**, 41–60 (2014)
19. Neurodiversity hiring: Global diversity and inclusion at microsoft. https://www.microsoft.com/en-us/diversity/inside-microsoft/cross-disability/neurodiversityhiring

20. Alozie, N.M., Dhamija, S.: Automated collaboration assessment using behavioral analytics. International Society of the Learning Sciences (2020)
21. Technologies U. https://unity.com/
22. Celebrating 20 years - global leader in eye tracking, April 2015. https://www.tobii.com/
23. Webcams. https://www.logitech.com/en-us/products/webcams.html
24. https://webrtc.org/
25. Iiyoshi, K., Tauseef, M., Gebremedhin, R., Gokhale, V., Eid, M.: Towards standardization of haptic handshake for tactile internet: a WEBRTC-based implementation. In: 2019 IEEE International Symposium on Haptic, Audio and Visual Environments and Games (HAVE), pp. 1–6 (2019)
26. Speech to text - audio to text translation: Microsoft azure. https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/#overview
27. Meier, A., Spada, H., Rummel, N.: A rating scheme for assessing the quality of computer-supported collaboration processes. Int. J. Comput.-Support. Collab. Learn. **2**(1), 63–86 (2007)