Social Performance Rating During Social Skills Training in Adults with Autism Spectrum Disorder and Schizophrenia

Kana Miyamoto [1] *miyamoto.kana.mk4@is.naist.jp*
Hiroki Tanaka [1] *hiroki-tan@is.naist.jp*
Jennifer Hamet Bagnou [2] *hamet@lisn.fr*
Elise Prigent [2] *elise.prigent@universite-paris-saclay.fr*
Céline Clavel [2] *celine.clavel@lisn.upsaclay.fr*
Jean-Claude Martin [2] *jean-claude.martin@lisn.upsaclay.fr*
Satoshi Nakamura [1] *s-nakamura@is.naist.jp*

[1] Nara Institute of Science and Technology, Ikoma, Japan
[2] Université Paris-Saclay, CNRS, Laboratoire Interdisciplinaire des Sciences du Numérique, 91400, Orsay, France

***Abstract*—Social communication skills are crucial factors influencing human social life. Quantifying the degree of social communication difficulties is necessary to understand developmental and neurological disorders and to create systems for automatic symptom screening and early intervention methods such as social skills training. Social skills training by human trainers is well established. Some automated social skills training systems have been proposed in the past. The previous social skills training system used a revised roleplay test to evaluate human social communication skills. However, this scale was evaluated by specialists such as psychiatrists and was not based on a standard internationally used scale. In this paper, we propose applying a social performance rating scale to social skills training data to measure social communication skills. We made a Japanese version of the social performance rating scale that can be rated without special qualifications and has English and French versions. In a previous study, we collected videos of interactions between trainers and adults with autism spectrum disorder, schizophrenia, and control participants during social skills training sessions. Two raters used the scale to annotate the collected data. The results found social performance rating scale showed good psychometric properties for assessing social communication skills. We found significantly greater gaze scores in adults with autism spectrum disorder than in adults with schizophrenia. There were differences between the ratings of different tasks in the adults with schizophrenia and control participants. These results suggest that the social performance rating scale can be a useful tool to assess social communication skills in different cultures and different pathologies. Possible future directions include using the social performance rating scale for assessing social behavior during interaction with a virtual agent.**

***Index Terms*—Social performance rating scale, social skills training, autism spectrum disorder, schizophrenia, annotations.**

I. Introduction

Social communication skills are essential in human social life. Experienced psychiatrists identify people with social communication difficulties such as autism spectrum disorder (ASD) and schizophrenia (Sz) through interviews based on diagnostic criteria, responses, and various neuropsychological tests [12]. To improve the identification accuracy of neurodevelopmental and neurological disorders such as ASD and Sz, discovering symptom-specific behavioral markers and phenotyping is crucial. Adults with ASD and Sz are characterized by social dysfunction [26]; in terms of deficits, there have been reports describing conversational gestures that were less closely synchronized with the co-occurring speech [7], frequency of eye contact [31], prosody [38], and delayed responses time [14]. For Sz, deficits in social cognition are a critical determinant of social functioning [8]. There are reports of people with ASD and Sz sharing several symptomatically similar characteristics such as flat or blunted affect (e.g., reduced eye contact) or alogia (e.g., impoverished speech) [10], [22]. Thus, quantifying the degree of social communication difficulties is necessary to understand the nature of ASD and Sz and to create systems for automatic symptom screening and early intervention methods such as social skills training (SST).

SST has been widely applied and adapted to help people who lack social communication skills. It is used in hospitals, employment support facilities, workplaces, schools, and other institutions. A human trainer generally conducts SST to promote appropriate social communication skills, strengthen the individual's social self-efficacy, and reduce social anxiety. The Bellack method [3], or step-by-step SST, is a well-structured and widely used approach that includes defining target skills, modeling, roleplay, feedback, and homework. This method defines the SST framework and its four basic skills: asking for requests, declining requests, telling positive feelings, and listening to others. However, it is cost-ineffective because those needing training must visit the place where the training is conducted (e.g., hospitals, employment support facilities). Accessibility is further limited due to the low number of professional trainers, especially in rural areas. Some researchers have therefore been conducting studies to build SST systems using socially interactive agents [28], [36].

Personalizing is an essential element for such SST systems. For example, Bellack et al. recommended using shorter and

more precise feedback for people with Sz who experience hallucinations [3]. Since they often hear voices distracting the SST, focusing is difficult. For ASD, feedback could be disengaging if trainers frequently mention exaggerated gestures and facial expressions since they are major symptomatic characteristics. Towards such personalized SST systems, qualitatively measuring the characteristics of ASD and Sz is imperative. Individual roleplay performance can be rated with revised roleplay test [30], [37], though there are no internationally validated scales. The test is further limited because it must be evaluated by experts such as psychiatrists and SST trainers.

This paper proposes applying the Social Performance Rating Scale (SPRS) to measure social communication skills. SPRS is a validated scale developed in English [11] and French [13]. The English version was applied to videos of a simulated dinner party for participants with a primary diagnosis of social phobia, with a primary diagnosis of an anxiety disorder, and control participants. The French version was applied to videos of collaborative games between neurotypical participants. This scale is well used in a variety of research [35]. We newly developed a Japanese version of the SPRS that can be used without special qualifications in SST. We collected four SST tasks by adults with ASD, with Sz, and control groups. Two raters annotated with the SPRS in SST. The goals of our study are 1) to validate our Japanese version of SPRS, 2) to apply SPRS to SST and examine correlations between SPRS and a revised roleplay test, 3) to investigate if SPRS can be applied to participants with ASD, with Sz, and control groups, and 4) to analyze the relations between SPRS and questionnaires.

In this paper, we contributed to developing and examining the following items.

- We newly developed a Japanese version of SPRS.
- We collected SST data and validated SPRS of Japanese.
- We examined SPRS and the revised roleplay test are related but not identical.
- We examined the social communication skills and the following items: differences of adults with ASD, with Sz, and control groups, task differences of SST, and questionnaire scores.

II. Materials and methods

A. Japanese Version of SPRS

The items of the English version of SPRS are provided in Table I. The SPRS is rated using a Likert scale [24] 1 to 5.

We newly developed a Japanese version of the SPRS. As suggested by Wild et al. [41] in their guide to good practice in translation and cultural adaptation, we first translated from English to Japanese and then from Japanese to English. This forward/backward translation method provides quality control to verify consistency between the translation and the original version. Inter Group Corp. (Osaka, Japan), a company specializing in language and cross-cultural adaptation, translated the SPRS at our request. This scale is available upon request to the authors.

B. SST Dataset

We used a Japanese human-to-human SST dataset collected in our previous study [29]. We collected SST data at Nara Institute of Science and Technology for the control group and at Nara Medical University for the clinical group. This experiment was approved by the ethics committee of Nara Institute of Science and Technology and Nara Medical University. We explained the procedure and purpose to the participants and obtained informed consent. The dataset includes data from 50 adults with the following characteristics: 16 in the ASD participants group, 15 in the Sz group, and 19 in the controls group (see Table II). We collected these psychiatric or neurodevelopmental clinical groups to investigate their social communication difficulties. We included these groups since the main SST clients in clinical facilities are ASD and Sz. We recruited the control group participants from a human recruitment agency. Two psychiatrists with SST experience joined this study as trainers (roleplay partners). The outpatient participants with ASD and Sz were not the first to meet the trainer. All participants performed SST of roleplay and feedback with the trainer. Four basic skills were performed: asking for requests, declining requests, telling positive feelings, and listening to others (asking, declining, telling, listening, in short). Each roleplay lasts one to three minutes. Some participants performed SST multiple times though our analysis was done on the first roleplay data.

We excluded participants who had undergone eye surgery or had a history of psychiatric hospitalization from the control group. For the clinical group, we excluded participants who scored less than 70 on the third edition of the Wechsler Adult Intelligence Scale (WAIS-III) intelligence quotient test [39]. We excluded participants with drug or alcohol addictions, personality disorders, or organic mental disorders. We set the sample size by taking the balance of the generality of the result and the practical workload for the data collection. The data collection period was from January 2020 to January 2021. Due to COVID-19 concerns, a transparent partition was placed between the participants and trainers. Two video cameras were placed behind each conversationalist to record the other individual at chest level from the participants' faces from the oblique angle. We also recorded a video and used two Kinect sensors from the side (see Fig. 1).

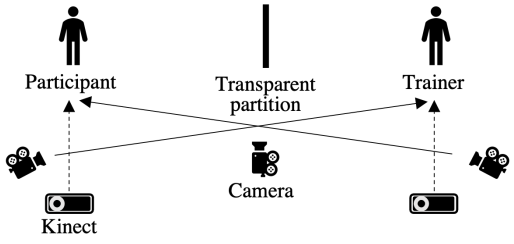

Fig. 1. SST data recording setup

TABLE I
SOCIAL PERFORMANCE RATING SCALE (SPRS) [11]

Item	Description
Gaze	(1) Very Poor: Participant completely avoids looking at the partner or stares continually. (5) Very Good: Participant keeps eye contact during the conversation, does not stare; shifts focus during pauses and conversation.
Voice Quality	(1) Very Poor: (a) Participant speaks in a flat, monotonous voice; or (b) speaks at a low volume or mumbles; or (c) speaks overly loudly; or has intrusive tone (harsh or unpleasant voice quality). (5) Very Good: Participant is warm and enthusiastic in verbal expression without sounding condescending or gushy.
Length	(1) Very Poor: Monosyllabic ('hmmm', 'yeah', 'OK') speech turns; or responses so long that partner must interrupt or cannot utter reply. (5) Very Good: At most times, participant's utterances are two or more sentences long. Participant acknowledges partner's remarks without taking over and monopolizing the conversation.
Discomfort	(1) Very High: Complete rigidity of arms, legs, or whole body. Constant leg movements or fidgeting with hands, hair, or clothing. Extremely stiff face or constant facial tics. Frequent nervous throat clearing, swallowing, or stuttering. Frequent inappropriate giggling or laughing. Look of extreme discomfort and desire to flee situation shown by 2 or more breaks in role. Participant does not pay attention to the roleplay tasks most of the time. (5) Very Low: Relaxed body posture and natural body movement. Participant laughs and smiles at appropriate times. S/he shows effective gesturing (to be distinguished from fidgeting). Participant focuses on the task all the time, does not appear at all uncomfortable, but at ease in situation.
Conversation Flow	(1) Very Poor: Participant makes few attempts to initiate the conversation. Even when prompted by the partner, participant cannot maintain the conversation. Participant uses almost no open-ended questions, or is intrusive in questions and shows no empathy. Participant does not attend to information provided by partner. (5) Very Good: Participant easily maintains the conversation and responds smoothly to pauses in the conversation, often by following up on previous information provided by the partner or providing free information about the self on a related topic. Participant introduces new topics fluidly and frequently uses open-ended questions. Participant shows genuine interest in the partner and follows up on the partner's remarks with warmth or enthusiasm.

TABLE II
PARTICIPANT DEMOGRAPHICS

	Control		ASD		Sz	
	Female	Male	Female	Male	Female	Male
N	9	10	6	10	8	7
Age	29.2 ±3.49	27.7 ±4.57	26.2 ±7.31	26.7 ±5.23	37.8 ±7.29	25.6 ±6.35

C. SPRS Ratings

Two third-party raters performed the annotations; all are non-experts in psychiatry, psychology, and SST. The raters watched the recorded videos from the oblique and side views for this evaluation. The raters evaluated all participants for the first roleplay of the four basic tasks during the SST. The raters watched the first three videos and discussed the SPRS content for evaluation practice, and then they evaluated the scores of the other videos without discussion. They were not told directly about the groups (ASD, Sz, controls), but it is possible that they may have noticed the groups by the background of the video because the control group and clinical group conducted SST in different locations. They were not given information about other questionnaire scales.

D. Questionnaire Scales

We assessed each participant using the following questionnaires for correlation analysis with the SPRS: the Facial Emotion Identification Test (FEIT) [15], Kikuchi's Scale of Social Skills: 18 items (KiSS-18) [21], Singelis' independent-interdependent Self-Construal Scale (Self-Construal Scale) [34], the second edition of the Social Responsiveness Scale (SRS-2) [4] (Japanese version [18]), and the Japanese version of the Brief Assessment of Cognition in Schizophrenia (BACS-J) [19]. In addition, we used the Autism Diagnostic Observation Schedule-second edition (ADOS-2) for the ASD group, and the Positive and Negative Syndrome Scale (PANSS) for the Sz group [1], [20]. The FEIT, KiSS-18, Self-Construal Scale, and SRS-2 were assessed by the participants themselves. The BACS-J, ADOS-2 and PANSS were assessed by psychiatrists. These scales are also used to evaluate SST systems [37]. Where possible, we obtained the total score and subscales of each questionnaire. Excluding the FEIT, the questionnaires were collected before the SST data collection.

FEIT, which assesses the emotional perception of facial emotions, includes facial images in a grayscale of 19 different people with one of six emotions: happiness, sadness, anger, surprise, fear, or disgust. We included this assessment because people with ASD and Sz struggle with social cognition in facial images [9]. KiSS-18, which measures social skill levels, is comprised of 18 questions based on six social skill categories. This metric comprehensively measures social skills. Self-Construal Scale, which consists of questions on a 7-point rating scale, was developed to measure how people view themselves in relation to others. SRS-2, composed of 65 questions, is an evaluation metric of the severity of social impairment. Although SRS-2 was initially designed to assess people with ASD, it can also differentiate among various social communication difficulties. Its effectiveness has been investigated with clinical and general population members [4].

For the participant evaluation during SST roleplay, we also adapted a revised roleplay test [30]. The scale is available only in Japanese. It includes items more related to psychopathology than the SPRS because the main target for this test is adults with Sz. For this test, third-party psychiatrists watch the recorded videos from the oblique and side views. The two psychiatrists used a Likert scale from 1 to 5 for eye contact,

body direction and distance, facial expression, voice variation, clarity, fluency, and social appropriateness. Since the required skills are situation dependent, the social appropriateness differs depending on each SST task. Let us explain examples of social appropriateness for Bellack's basic tasks. Listening to others, which determines whether the participants paid attention to the interlocutor, includes nodding, back-channels, and other nonverbal behaviors (e.g., eye contact, smiling). For expressing positive feelings, social appropriateness involves expressing attention to the interlocutor's responses and the suitability of the participant's speech content. For making requests, social appropriateness assesses whether they explained the details of their request, including what kind of help they needed. It also includes whether they listened to the interlocutor. For the declining, social appropriateness concerns whether they expressed remorse and appropriate reasons for their refusal. It also includes whether they proposed alternatives to the requests (e.g., I'm sorry, but I propose to do it next time), which is essential for the situation.

E. Statistical Analysis

The analyses were performed to investigate whether the SPRS could be used to evaluate social communication skills during SST. We analyzed the inter-rater agreement, items' characteristics, construct validity, and internal consistency. To check the inter-rater agreement, we calculated the intraclass correlation coefficient (ICC, 2k) [33]. We confirmed the intraclass correlation coefficients were more than 0.509, as reported in Section III-A. We averaged the SPRS scores for the two raters for our analysis. We calculated the Spearman's correlation coefficients to examine the homogeneity of the SPRS and the revised roleplay test. We assessed the factor structure of the SPRS by exploratory factor analysis with oblimin rotation to confirm the construct validity [5]. We examined Bartlett's Test of Sphericity ($p < 0.001$) and the Kaiser–Meyer–Olkin Test for Sampling Adequacy of 0.881 to apply the exploratory factorial analyses [2]. The internal consistency was assessed by calculating Cronbach alpha coefficient [6] and McDonald omega [27].

We also analyzed the total of SPRS items in relation to the three groups, four tasks of SST, and questionnaires. In comparing the control, ASD and Sz groups, we used the nonparametric Kruskal–Wallis test. Post-hoc analysis was conducted using the Mann-Whitney U test with the Bonferroni correction. In comparing the asking, declining, telling, and listening tasks, we used the Wilcoxon signed-rank test with Bonferroni correction. The Spearman's correlation coefficients were calculated between the total SPRS items and questionnaire scores for the control, ASD, and Sz groups. We used the JASP (Jeffreys's Amazing Statistics Program) for the statistical analysis [16].

III. Results

A. Inter-rater Agreement

We measured the inter-rater agreement by calculating the intraclass correlation coefficient (ICC, 2k) for validation of the SPRS annotated by the two raters. We opted for the random two-way ICC method with the mean as the unit of evaluation shown in Table III. The intraclass correlation coefficient ranged from 0 to 1 and all items had reliability values between 0.509 and 0.804. The intraclass correlation coefficients were above 0.5, which indicates more than moderate reliability [23]. The score for each SPRS item was averaged for the two raters' annotations in subsequent analyses.

TABLE III
Inter-rater reliability for the SPRS

Item	ICC 2K
Gaze	0.509
Voice Quality	0.795
Length	0.777
Discomfort	0.790
Conversation Flow	0.804

TABLE IV
Factor loading for SPRS score (U: Uniqueness, SD: Standard Deviation, S: Skewness, K: Kurtosis)

Item	Factor 1	U	Mean	SD	S	K
Gaze	0.881	0.224	3.21	1.00	-0.351	-0.726
Voice Quality	0.914	0.164	2.93	1.18	0.168	-0.975
Length	0.928	0.139	3.27	1.10	-0.247	-0.819
Discomfort	0.874	0.236	3.04	1.02	-0.074	-0.659
Conversation Flow	0.932	0.132	3.15	1.07	-0.192	-0.585

B. Item Characteristics

The Spearman's correlation performed between the SPRS items and revised roleplay test items shows significant results (all $p<0.001$) and positive relationships (Fig. 2). The correlation between any of them indicated more than 0.6.

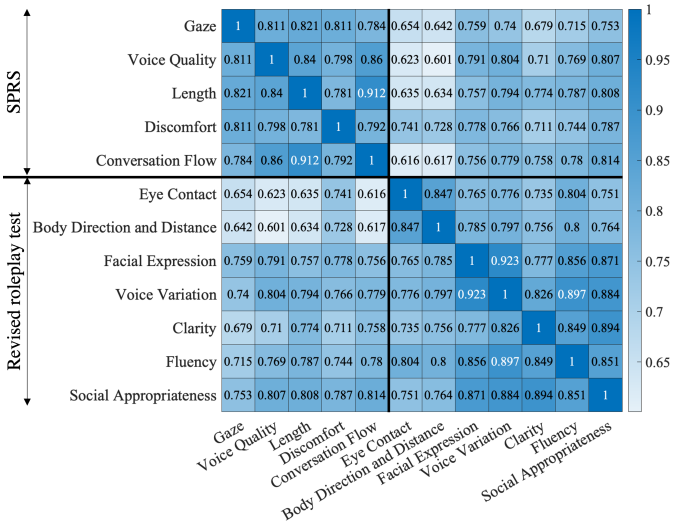

Fig. 2. Spearman's inter-item correlation for the SPRS

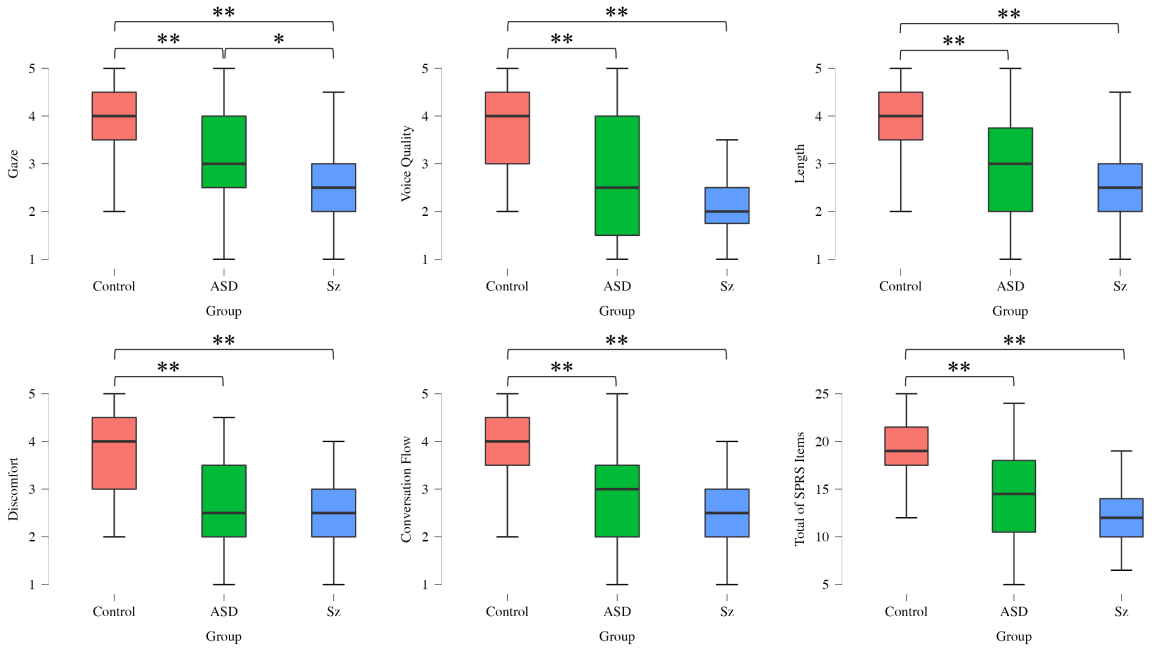

Fig. 3. Boxplot of SPRS items for control, ASD, and Sz groups (*p=0.005, **p<0.001)

C. Construct Validity

Exploratory factorial analyses were performed on the five items of SPRS with an oblique rotation (oblimin) shown in Table IV. All items have a factor load between 0.874 and 0.932. The results show that the first factor explained 82.1% of the variance.

D. Internal Consistency

We created an overall performance measure by summing the scores of the five SPRS items. This measure demonstrates that Cronbach alpha was 0.957 and McDonald omega was 0.958.

E. Kruskal–Wallis Test for Three Groups

Boxplots of the five items of SPRS and the total of the items for the control, ASD, and Sz groups are shown in Fig. 3. We used the Kruskal–Wallis test to find any differences in the SPRS items between the three groups. There were significant differences in all items and the total of the SPRS items ($p<0.001$). Post-hoc analysis was conducted using the Mann-Whitney U test with Bonferroni correction. We found significant differences in gaze, voice quality, length, discomfort, conversation flow, and the total of SPRS items for control-ASD and control-Sz ($p<0.001$). In ASD-Sz, we found significantly greater gaze scores in the ASD than Sz groups ($p=0.005$, Rank-Biserial Correlation=0.290).

F. Wilcoxon Signed-rank Test for Four Tasks

We performed the Wilcoxon signed-rank test with Bonferroni correction for each group to find significant differences in the four tasks. Table V shows the items that differed significantly among the four tasks. In the control group, we found significant differences in all items except gaze in asking-declining tasks and telling-declining tasks, and there were significant differences in the length of the asking-listening and telling-listening tasks. The declining task had lower scores than the asking and telling tasks for the control group. The length item of the listening task also had a low score compared to asking and telling tasks. We did not find significant differences between the four tasks in the ASD group. In the Sz group, we found significant differences in the length of telling-listening tasks. For length, the listening task had a lower score than the telling task.

G. Spearman's Correlation for Questionnaire Scores

Spearman's correlation coefficients were calculated between the total SPRS items and questionnaire scores for the control, ASD, and Sz groups. The correlation coefficients are shown in Table VI. In the control group, KiSS-18 and the Self-Construal Scale had significant correlations with the total SPRS items. In the ASD group, the following had significant correlations: KiSS-18, Self-Construal Scale, SRS-2, BACS-J, and ADOS-2. In the Sz group, the following had significant correlations: FEIT, BACS-J, and PANSS.

IV. Discussion

A. SPRS for Evaluation of SST Roleplay

In this paper, we propose to evaluate SST using the SPRS. We calculated the inter-rater reliability between the two raters' evaluated SPRS scores and found that the values were above 0.509 for any SPRS item. Our results show that the reliability was moderate to good [23]. Hamet Bagnou et al. calculated the reliability among SPRS raters during collaborative tasks [13]. They showed that the reliability of gaze was the highest, but our results showed the lowest values. One reason could be the difference in conversation content. In our study, we

TABLE V
DIFFERENCES IN SPRS ITEMS FOR EACH TASK (BOLD MEANS P <0.0083). [] INDICATES SIGNIFICANTLY HIGHER SCORED TASKS. EFFECT SIZE IS RANK-BISERIAL CORRELATION.

Item	Combination		Control		ASD		Sz	
			p	Effect size	p	Effect size	p	Effect size
Gaze	Asking	Declining	0.066	0.603	0.078	-0.591	0.597	-0.200
		Telling	0.468	-0.289	0.356	-0.295	0.820	-0.091
		Listening	0.477	0.219	0.914	-0.044	0.160	0.485
	Declining	Telling	0.013	-0.743	0.401	0.309	0.968	0.026
		Listening	0.154	-0.386	0.174	0.491	0.104	0.526
	Telling	Listening	0.183	0.400	0.240	0.500	0.234	0.397
Voice Quality	Asking	Declining	**<0.001** [Asking]	0.963	0.588	-0.250	0.608	0.200
		Telling	0.450	0.242	0.454	-0.289	0.224	-0.385
		Listening	0.131	0.515	0.813	0.091	0.112	0.564
	Declining	Telling	**0.001** [Telling]	-0.950	0.968	-0.026	0.119	-0.513
		Listening	0.039	-0.600	0.630	0.200	0.491	0.255
	Telling	Listening	0.283	0.341	0.244	0.418	0.026	0.758
Length	Asking	Declining	**0.001** [Asking]	0.933	0.523	0.198	0.615	0.222
		Telling	0.627	-0.167	0.100	0.561	0.427	-0.291
		Listening	**0.008** [Asking]	0.824	0.150	0.527	0.035	0.727
	Declining	Telling	**<0.001** [Telling]	-1.000	0.299	0.364	0.299	-0.364
		Listening	0.560	-0.175	0.236	0.350	0.060	0.652
	Telling	Listening	**0.003** [Telling]	0.912	0.472	0.258	**0.005** [Telling]	0.923
Discomfort	Asking	Declining	**0.001** [Asking]	0.933	0.876	-0.073	0.829	-0.111
		Telling	0.902	-0.039	0.809	0.111	0.903	0.067
		Listening	0.493	0.220	0.595	-0.286	0.294	0.382
	Declining	Telling	**0.001** [Telling]	-0.912	0.671	0.194	0.672	0.238
		Listening	0.022	-0.647	0.615	-0.222	0.245	0.385
	Telling	Listening	0.412	0.248	0.492	-0.267	0.320	0.417
Conversation flow	Asking	Declining	**0.004** [Asking]	0.797	0.592	-0.179	0.515	0.200
		Telling	0.666	-0.164	0.202	0.423	0.151	-0.474
		Listening	0.033	0.603	0.500	0.242	0.151	0.500
	Declining	Telling	**<0.001** [Telling]	-0.949	0.202	0.423	0.149	-0.474
		Listening	0.247	-0.342	0.194	0.423	0.365	0.318
	Telling	Listening	0.016	0.700	1.000	0.010	0.026	0.731
Total of SPRS items	Asking	Declining	**<0.001** [Asking]	0.916	0.406	-0.250	0.959	0.036
		Telling	0.630	-0.135	0.614	0.162	0.310	-0.330
		Listening	0.067	0.510	0.484	0.206	0.033	0.681
	Declining	Telling	**<0.001** [Telling]	-0.979	0.379	0.276	0.394	-0.267
		Listening	0.085	-0.468	0.267	0.333	0.146	0.487
	Telling	Listening	0.022	0.614	0.531	0.192	0.017	0.758

applied SPRS to four SST tasks. The annotators needed to evaluate social communication skills based on the content of the SST tasks. The evaluation of our tasks may have been more complex than the collaborative tasks. There are slight differences between their study and ours but the results are almost identical, with a reliability of 0.5 or higher in both studies.

The Spearman's correlation between the SPRS items and revised roleplay test items was above 0.6. The factor analysis for SPRS showed that the first factor explained 82.1% of the variance. The Cronbach alpha was 0.957 and McDonald omega was 0.958. These results indicate that SPRS shows excellent internal consistency and is related to the revised roleplay test but not identical.

B. Differences in SPRS for Groups and Tasks

We compared the SPRS scores by group and task. The control group differed significantly from the ASD and Sz groups on any item of the SPRS. Several studies have shown that the control group shows higher social communication skills than the ASD and Sz groups, and we showed similar results. The ASD group had a higher gaze score than Sz group. We consider it related that the eye movement abnormalities are severe in the Sz group compared to the ASD group [32]. We conducted analyses of the differences in the SPRS scores between tasks for each group. In the control group, the declining task had lower scores than the asking and telling tasks, except for gaze. Declining is the most difficult task of the basic SST tasks, which may have contributed to the lower scores [3]. The listening task had lower scores than the asking and telling tasks in length. In the Sz group, the listening task also scored lower than the telling task in length. The control group met the trainer for the first time in the data recording. It is possible that some control participants were not interested in the topics provided by the trainer in the listening task. The ASD and Sz groups had met with the trainer several times. However, the Sz group had low scores. Some participants with Sz might have seemed disinterested in the conversation [3], and the raters might have annotated the low scores.

C. Correlation Between SPRS and Questionnaire Scores

The control group had significant positive correlations between the SPRS and the KiSS-18, and the Self-Construal Scale. Independence of the Self-Construal Scale measures is

TABLE VI
SPEARMAN'S CORRELATION BETWEEN THE TOTAL OF SPRS ITEMS AND QUESTIONNAIRE SCORES (*P<0.05, **P<0.01, ***P<0.001)

Item		Control	ASD	Sz
FEIT		0.124	0.146	0.548***
KiSS-18	Total Score	0.278*	0.307*	0.005
	Basic Skills	0.086	0.438***	-0.159
	Advanced Skills	0.467***	0.407***	0.201
	Emotional Management Skills	0.124	0.206	-0.093
	Offence Management Skills	0.181	0.407***	0.086
	Stress Management Skills	0.348**	-0.096	0.155
	Planning Skills	-0.136	-0.035	0.081
Self-Construal Scale	Total Score	0.298**	0.365**	0.067
	Independent-Interdependence	-0.195	-0.347**	-0.240
	Independent	0.064	-0.034	-0.065
	Interdependence	0.489***	0.327**	0.220
SRS-2		-0.092	-0.817***	-
BACS-J	Verbal Memory	-	0.026	0.338**
	Working Memory	-	-0.074	0.193
	Motor Speed	-	0.067	0.548***
	Verbal Fluency	-	0.173	0.455***
	Semantic Fluency	-	-0.054	0.380**
	Symbol Coding	-	0.194	0.465***
	Attention	-	-0.215	0.565***
	Executive Function	-	-0.282*	0.225
ADOS-2	Language and Communication	-	-0.196	-
	Reciprocal Social Interaction	-	-0.731***	-
	Social Affect Total	-	-0.560***	-
	Restricted and Repetitive Behaviors	-	-0.234	-
	Social Affect and Restricted and Repetitive Behavior Total	-	-0.472***	-
PANSS	Total Score	-	-	-0.644***
	Positive	-	-	-0.282*
	Negative	-	-	-0.643***
	General Psychopathology	-	-	-0.555***

characterized by autonomy and distinction from others [17] and is embedded within the culture [25]. In cultures with strong interdependence, harmony is an important concern, and assertiveness tends to be discouraged. It is predominant in the East, including Japan, and may be related to social communication skills in Japan.

In the ASD group, there were significant positive or negative correlations between the SPRS and the KiSS-18 and Self-Construal Scale, as well as the control group. We found significant negative correlations in SRS-2, BACS-J, and ADOS-2. Our results suggest a relationship between the questionnaire for ASD and the social communication skills of the ASD group.

The Sz group had significant positive or negative correlations between the SPRS and FEIT, BACS-J, and PANSS. Emotional awareness and communication skills are related [40], and we surmise that our results show a significant difference in FEIT. The questionnaire for Sz also associated the social communication skills of Sz and the results of the Sz group.

V. CONCLUSIONS

We applied and validated the SPRS of Japanese to SST data to measure individual social communication skills. We analyzed four SST tasks of adults with ASD and with Sz and control participants. The results found significant differences between the control-ASD and control-Sz in all SPRS items. We found significantly greater gaze scores in the ASD than in the Sz groups. There were task differences in the Sz and control groups. Our results suggest that SPRS can be a useful tool to assess social communication skills in different cultures and different pathologies and can evaluate SST effectiveness as well as the revised roleplay test. Possible future directions include using the social performance rating scale for assessing social behavior during interaction with a virtual agent.

ACKNOWLEDGMENT

This work was supported by JST, CREST Grant Number JPMJCR19A5 and the establishment of university fellowships towards the creation of science technology innovation Grant Number JPMJFS2137.

ETHICAL IMPACT STATEMENT

All SST data collection processes were approved by ethical committees at Nara Institute of Science and Technology and Nara Medical University. At the beginning of the data collection, we explained the procedure and purpose to the participants and obtained informed consent. Our findings are limited in terms of the small number of participants and the mild cases of ASD and Sz participants who can verbally communicate. Further validations are needed to examine the generalizability of the results.

References

[1] Jojanneke A Bastiaansen, Harma Meffert, Simone Hein, Petra Huizinga, Cees Ketelaars, Marieke Pijnenborg, Arnold Bartels, Ruud Minderaa, Christian Keysers, and Annelies De Bildt. Diagnosing autism spectrum disorders in adults: the use of autism diagnostic observation schedule (ados) module 4. *Journal of autism and developmental disorders*, 41:1256–1266, 2011.

[2] Amy S Beavers, John W Lounsbury, Jennifer K Richards, Schuyler W Huck, Gary J Skolits, and Shelley L Esquivel. Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research, and Evaluation*, 18(1):6, 2013.

[3] Alan S Bellack, Kim T Mueser, Susan Gingerich, and Julie Agresta. *Social skills training for schizophrenia: A step-by-step guide*. Guilford Publications, 2013.

[4] John N Constantino and Christian P Gruber. *Social responsiveness scale: SRS-2*. Western psychological services Torrance, CA, 2012.

[5] Anna B Costello and Jason Osborne. Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical assessment, research, and evaluation*, 10(1):7, 2005.

[6] Lee J Cronbach. Coefficient alpha and the internal structure of tests. *psychometrika*, 16(3):297–334, 1951.

[7] Ashley de Marchena and Inge-Marie Eigsti. Conversational gestures in autism spectrum disorders: Asynchrony but not decreased frequency. *Autism research*, 3(6):311–322, 2010.

[8] David Dodell-Feder, Laura M Tully, and Christine I Hooker. Social impairment in schizophrenia: new approaches for treating a persistent problem. *Current opinion in psychiatry*, 28(3):236, 2015.

[9] João Miguel Fernandes, Rute Cajão, Ricardo Lopes, Rita Jerónimo, and J Bernardo Barahona-Corrêa. Social cognition in schizophrenia and autism spectrum disorders: a systematic review and meta-analysis of direct comparisons. *Frontiers in psychiatry*, 9:504, 2018.

[10] Jennifer H Foss-Feig, James C McPartland, Alan Anticevic, and Julie Wolf. Re-conceptualizing asd within a dimensional framework: Positive, negative, and cognitive feature clusters. *Journal of autism and developmental disorders*, 46:342–351, 2016.

[11] Thomas Fydrich, Dianne L Chambless, Kevin J Perry, Friederike Buergener, and Maria B Beazley. Behavioral assessment of social performance: A rating system for social phobia. *Behaviour research and therapy*, 36(10):995–1010, 1998.

[12] Martin Guha. Diagnostic and statistical manual of mental disorders: Dsm-5. *Reference Reviews*, 2014.

[13] Jennifer Hamet Bagnou, Jean-Claude Martin, Élise Prigent, and Céline Clavel. Adaptation and validation of two annotation scales for assessing social skills in a corpus of multimodal collaborative interactions. *Frontiers in Psychology*, page 6647, 2022.

[14] Peter A Heeman, Rebecca Lunsford, Ethan Selfridge, Lois M Black, and Jan Van Santen. Autism and interactional aspects of dialogue. In *Proceedings of the SIGDIAL 2010 Conference*, pages 249–252, 2010.

[15] William P Horan, Robert S Kern, Karina Shokat-Fadai, Mark J Sergi, Jonathan K Wynn, and Michael F Green. Social cognitive skills training in schizophrenia: an initial efficacy study of stabilized outpatients. *Schizophrenia research*, 107(1):47–54, 2009.

[16] JASP Team. JASP (Version 0.17)[Computer software], 2023.

[17] Chester Kam, Xiaolin Zhou, Xuan Zhang, and Man Yee Ho. Examining the dimensionality of self-construals and individualistic–collectivistic values with random intercept item factor analysis. *Personality and Individual Differences*, 53(6):727–733, 2012.

[18] Yoko Kamio, Naoko Inada, Aiko Moriwaki, Miho Kuroda, Tomonori Koyama, Hiromi Tsujii, Yuki Kawakubo, Hitoshi Kuwabara, Kenji J Tsuchiya, Yota Uno, et al. Quantitative autistic traits ascertained in a national survey of 22 529 j apanese schoolchildren. *Acta Psychiatrica Scandinavica*, 128(1):45–53, 2013.

[19] Yasuhiro Kaneda, Tomiki Sumiyoshi, Richard Keefe, Yasuhito Ishimoto, Shusuke Numata, and Tetsuro Ohmori. Brief assessment of cognition in schizophrenia: validation of the japanese version. *Psychiatry and Clinical Neurosciences*, 61(6):602–609, 2007.

[20] SR Kay, A Fiszbein, and LA Opler. The positive and negative syndrome scale (panss) for schizophrenia. schizophr. bull. 13 (2), 261-276. 1987.

[21] A Kikuchi. Measure social skill: handbook of kiss-18 (110-115), 2007.

[22] M Mary Konstantareas and Terri Hewitt. Autistic disorder and schizophrenia: Diagnostic overlaps. *Journal of autism and developmental disorders*, 31:19–28, 2001.

[23] Terry K Koo and Mae Y Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2):155–163, 2016.

[24] Rensis Likert. A technique for the measurement of attitudes. *Archives of psychology*, 1932.

[25] Hazel R Markus and Shinobu Kitayama. Culture and the self: Implications for cognition, emotion, and motivation. *Psychological review*, 98(2):224, 1991.

[26] Steven Marwaha and Sonia Johnson. Schizophrenia and employment: a review. *Social psychiatry and psychiatric epidemiology*, 39:337–349, 2004.

[27] Roderick P McDonald. *Test theory: A unified treatment*. psychology press, 2013.

[28] Jacqueline Nadel, Ouriel Grynszpan, and Jean-Claude Martin. Autism and socially interactive agents. In *The Handbook on Socially Interactive Agents: 20 years of Research on Embodied Conversational Agents, Intelligent Virtual Agents, and Social Robotics Volume 2: Interactivity, Platforms, Application*, pages 437–462. 2022.

[29] Takeshi Saga, Hiroki Tanaka, Yasuhiro Matsuda, Tsubasa Morimoto, Mitsuhiro Uratani, Kosuke Okazaki, Yuichiro Fujimoto, and Satoshi Nakamura. Automatic evaluation-feedback system for automated social skills training. *Scientific Reports*, 13(1):6856, 2023.

[30] Takashi Sasaki. The reliability and validity of the revised role play test for assessment of social skills in schizophrenia. *Clinical Psychiatry*, 48(11):1191–1198, 2006.

[31] Atsushi Senju and Mark H Johnson. Atypical eye contact in autism: models, mechanisms and development. *Neuroscience & Biobehavioral Reviews*, 33(8):1204–1214, 2009.

[32] Tomoko Shiino, Kenichiro Miura, Michiko Fujimoto, Noriko Kudo, Hidenaga Yamamori, Yuka Yasuda, Manabu Ikeda, and Ryota Hashimoto. Comparison of eye movements in schizophrenia and autism spectrum disorder. *Neuropsychopharmacology Reports*, 40(1):92–95, 2020.

[33] Patrick E Shrout and Joseph L Fleiss. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2):420, 1979.

[34] Theodore M Singelis and William F Sharkey. Culture, self-construal, and embarrassability. *Journal of cross-cultural psychology*, 26(6):622–644, 1995.

[35] Stephan Stevens, Ruth Cooper, Trisha Bantin, Christiane Hermann, and Alexander L Gerlach. Feeling safe but appearing anxious: Differential effects of alcohol on anxiety and social performance in individuals with social anxiety disorder. *Behaviour research and therapy*, 94:9–18, 2017.

[36] Hiroki Tanaka, Hidemi Iwasaka, Yasuhiro Matsuda, Kosuke Okazaki, and Satoshi Nakamura. Analyzing self-efficacy and summary feedback in automated social skills training. *IEEE Open Journal of Engineering in Medicine and Biology*, 2:65–70, 2021.

[37] Hiroki Tanaka, Takeshi Saga, Kota Iwauchi, Masato Honda, Tsubasa Morimoto, Yasuhiro Matsuda, Mitsuhiro Uratani, Kosuke Okazaki, and Satoshi Nakamura. The validation of automated social skills training in members of the general population over 4 weeks: Comparative study. *JMIR Formative Research*, 2023.

[38] Hiroki Tanaka, Sakriani Sakti, Graham Neubig, Tomoki Toda, and Satoshi Nakamura. Linguistic and acoustic features for automatic identification of autism spectrum disorders in children's narrative. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality*, pages 88–96, 2014.

[39] David S Tulsky, Nancy D Chiaravalloti, Barton W Palmer, and Gordon J Chelune. The wechsler memory scale: a new perspective. In *Clinical interpretation of the WAIS-III and WMS-III*, pages 93–139. Elsevier, 2003.

[40] Gülten Uzun and Neslihan Lok. Effect of emotional awareness skills training on emotional awareness and communication skills in patients with schizophrenia: A randomized controlled trial. *Archives of Psychiatric Nursing*, 38:14–20, 2022.

[41] Diane Wild, Alyson Grove, Mona Martin, Sonya Eremenco, Sandra McElroy, Aneesa Verjee-Lorenz, and Pennifer Erikson. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (pro) measures: report of the ispor task force for translation and cultural adaptation. *Value in health*, 8(2):94–104, 2005.