Measuring group social interactions following acquired brain injury: an inter-rater reliability evaluation

Susan Howell, PhD,[a] Rosemary Varley, PhD,[a] Emma Louise Sinnott, MSc,[c] Tim Pring, PhD,[b] Suzanne, Beeke, PhD[a]

[a] Division of Psychology and Language Sciences, University College London

[b] Division of Language and Communication Science, City, University of London

[c] Central London Community Healthcare Trust

Corresponding author: Susan Howell, Department of Language and Cognition, University College London, Chandler House, 2 Wakefield Street, London WC1N 1PF

Tel: +44 207 679 4001

Email: s.howell.12@ucl.ac.uk

Abstract

Background

Social communication impairments following acquired brain injury (ABI) are well-documented. There is evidence that group interventions are beneficial but research into validated instruments to measure group outcomes is a new field of investigation.

Aims

This study reports on the inter-rater reliability of three established social communication measures for use with group interaction data: the Profile of Pragmatic Impairment in Communication (PPIC), the Behaviorally Referenced Rating System of Intermediate Social Skills (BRISS-R), the Adapted Measure of Participation in Conversation (MPC). Inter-rater reliability of the Interactional Network Tool (INT), a new digital tool designed for group interactional behaviours, is also evaluated.

Method

Thirty one video samples of ABI group interactions were independently rated by two rater pairs using the four outcome measures. Inter-rater reliability was calculated using intra-class correlations (ICC).

Results

ICC estimates and their 95% confidence intervals were calculated for the different measures. The measures showed differential sensitivity. Rater agreement on the MPC interaction (ICC=0.77) and transaction (ICC=0.74) scales was moderate to good. The INT initiation frequencies (ICC=0.83) were moderate to excellent and the INT response frequencies (ICC=0.69) were poor to good. Poor to moderate reliability was achieved on the BRISS-R PCSS (ICC=0.49) and PDBS (ICC=0.50) scale and PPIC findings were moderate but showed presence of skew.

Conclusion

Acceptable reliability was achieved on two measures of participation (MPC and INT). The INT shows promise as a new method to characterise interactions and detect change in group communication behaviour.

Key words: brain injury, rehabilitation, rating scales, outcome measurement, social networks, group

Background and Aims

Changes in social competence and their adverse impact on relationship-building and social acceptance are commonly reported following acquired brain injury (ABI) (Douglas et al., 2019). Behaviours with the potential to disrupt the conversational interchange or infringe social norms following ABI are wide-ranging as a result of the complex interplay of cognitive, linguistic, social, and emotional factors that impact communication competence (MacDonald 2017). Intervention to address disordered social communication is a practice recommendation in brain injury rehabilitation, and treatments typically encompass both individual and group approaches (Togher, Wiseman-Hakes et al., 2014).

Findings from systematic reviews published in the past 10 years show evidence of effectiveness for interventions for social communication skills. Cicerone et al. (2011) cited evidence from four group studies to conclude that there was sufficient evidence for interventions to be delivered as a practice standard. Finch et al. (2016) concluded from a review of 15 group and individual studies that interventions for impaired social communication skills were generally beneficial, but the evidence for group interventions sensitive to the communication context was strongest. Notably, the group intervention studies cited in these reviews trained skills in group settings and measured change in dyadic interactions (typically between the person with ABI and a neurotypical communication partner).

Post-injury, effective social and interpersonal skills are pre-requisite to returning to previous roles at home and in the wider community, and to the maintenance of

previous friendships and the formation of new ones. Social participation in everyday life entails interaction in group settings, and requires the ability to independently and simultaneously engage with multiple communication partners. The behavioural characteristics following ABI that impede successful social participation encompass a wide distribution, from communicative excess to insufficiency (Hartley 1995). In addition to the cognitive and language impairments that preclude effective information sharing, poor behavioural control and problems recognising and accurately interpreting social cues can prevent effective participation in interaction. Therapy groups and social interaction groups are a means to practise and develop these skills following injury, and are commonplace in post-acute and rehabilitation settings, and social and leisure activity groups are routinely provided in residential and day centre settings (Hammond et al., 2015). Given this landscape, the absence of validated instruments designed to measure group interaction is surprising.

The current investigation identified three established social communication measures developed to assess and evaluate social communication skills in TBI: the Profile of Pragmatic Impairment in Communication (PPIC) (Linscott, Knight and Godfrey, 1996), two subscales from the Behaviorally Referenced Rating System of Intermediate Social Skills (BRISS-R) (Farrell, Rabinowitz, Wallander and Curran, 1985) and the Adapted Measure of Participation in Conversation (MPC) (Togher et al. 2010). Each has previously been employed to measure social communication in controlled trials of group interventions following ABI. In each case, the measures have been designed to evaluate dyadic interaction data. This investigation tested their reliability with group interaction data. The evaluation preceded a pilot study comparing the effects of a peer-led social communication intervention versus usual

care (Howell et al., 2020). One new measure, the Interactional Network Tool (INT) (Howell, 2018) was also evaluated. This measure was developed specifically to evaluate conversation behaviour in an ABI group.

The three established measures were designed as assessment and profiling tools to evaluate social communication behaviour from filmed interactions. The PPIC and BRISS-R are both observational rating scales of social behaviours. The PPIC was designed to assess social communication behaviour and aspects of language content. These are evaluated across 10 subscales (including conversation content, clarity, participation and social style), using an established model of pragmatic communication (Grice, 1975). The BRISS-R is comprised of six social behaviour rating scales. The Personal Conversational Style Scale (PCSS) and the Person Directed Behaviour Scale (PDBS) measure social communication behaviour in the context of others. These sub-scales evaluate social manners, use of humour, degree of self-disclosure, the use of reinforcers, self-centred behaviour and partner directed behaviour, and are therefore sensitive to the loss of higher-level social judgement and awareness that frequently follows ABI. Another observational tool, the MPC, also measures skills in the context of others. Social participation is evaluated using two sub-scales. The interaction sub-scale, measures the ability of the person with ABI to participate verbally and non-verbally in a conversation. The transaction sub-scale measures the ability to share and request information.

In addition to their role as profiling tools, these instruments were employed to measure outcomes in the parallel group studies cited in recent systematic reviews. They were also evaluated in a state-of-the-art review of social communication

measurement tools (Sohlberg et al., 2019) and each was rated positively for published reliability data. The PPIC has previously been used to evaluate the efficacy of a manualised social communication skills training programme (Dahlberg et al., 2007; Braden et al., 2010) and a metacognitive strategy intervention (Finch et al., 2017). Linscott et al. (1996) previously reported good inter-rater reliability (ICC= 0.75-0.88) for the PPIC. McDonald et al. (2008) employed the PCSS and the PDBS from the BRISS-R to evaluate their manualised social skills intervention. Flanagan, McDonald and Togher (1995) reported inter-rater reliability as good to excellent for this measure, but McDonald, Tate et al. (2008) reported more variable findings (PDBS findings: ICC=0.43-0.83, PCSS findings 0.097-0.682). Togher et al. (2013) employed the MPC to evaluate a training programme for everyday communication partners. Excellent inter-rater reliability (ICC=0.84-0.89) was reported in Togher et al. (2010).

The suitability of an assessment tool to measure intervention outcomes depends on how well the construct under evaluation maps on to the behaviour targeted by the intervention (Tate, 2019). Published reliability data for the established measures are reported from dyadic conversation ratings, and the ability of these instruments to reliably measure change in group interaction cannot be assumed. In a group conversation, individuals are required to monitor and contribute to exchanges between multiple interactors. Group conversation can take an unpredictable and fluctuating course that necessitates initiating and responding adaptively to others. The ability of the existing measures using rating scale data to assess adaptive behaviour cannot be guaranteed. Further, reliability was tested on samples of a person with ABI in conversation with a neuro-typical conversation partner. Group

initiatives in post-acute, rehabilitation and community settings comprise ABI peers (with and without support from neurotypical interactors), and the ability of these measurement tools to sample behaviours relevant to conversations between peers requires investigation.

Against this background the Interactional Network Tool (INT) was developed to directly assess peer interactions and groups involving neurotypical conversation partners. In contrast to the existing measures, the INT uses relational data in the form of behavioural frequencies. It draws on social network theories and analysis methods (Scott, 2017; Wasserman and Faust, 1994) and the proposition that individuals can be influenced to modify their interactional behaviour by their social network (Bandura, 1971; Lave and Wenger, 1991; Vygotsky, 1978). It is a methodological approach that enables group interaction patterns to be visualised. Its theoretical basis is in discourse analysis models using initiation and response categories to analyse interaction (Coulthard, 1984; Eggins and Slade, 1997). These have previously been tested in ABI studies as a method to evaluate interaction effectiveness (Coehlo et al., 1991; Coehlo et al., 2002). The INT also has a theoretical foundation in models of social communication skills (Trower et al., 1978), incorporating verbal and non-verbal behaviours as important contributors to the interaction flow.

This study reports on the inter-rater reliability of the BRISS-R, PPIC, MPC and the INT to evaluate ABI group interactions. Samples were drawn from an investigation into the feasibility of a peer-led intervention for social communication skills. This evaluation preceded selection of the primary outcome measures for a subsequent

trial to investigate whether a peer-led intervention is more effective than a staff-led activity group (ClinicalTrials.gov PRS: NCT02211339). The aim of this report is to:

- evaluate the inter-rater reliability of established social communication measures for use with group interaction data
- report on the inter-rater reliability of the INT, the new digital tool specifically designed to evaluate group interactional behaviours

Method

Twelve participants (seven males and five females) with severe TBI or severe ABI with similar cognitive presentation to TBI took part in the investigation. They were recruited from a specialist residential neurorehabilitation centre. All participants had a social communication impairment (judged by the treating clinical team), were aged between 18 and 65 years and of white ethnicity. TBI severity was measured by post-traumatic amnesia (PTA) duration exceeding 24 hours (or other neurological evidence e.g. persisting neurological signs). ABI severity was measured by a Glasgow Coma Scale (GCS) (Teasdale and Jennett, 1974) score of less than 9 or other persisting neurological signs. Demographic and injury-related data are presented in Table 1.

Table 1: Demographic data

Participant	Age (years)	Male/ female	Education (years)	TPO (years)	Severity	Aetiology
P1	60	M	13	41	Severe	TBI: RTA and falls
P2	39	M	18	0.8	Severe	ABI: ICH
P3	63	M	13	25	Severe	ABI: RTA and falls
P4	42	F	13	7	Severe	ABI: Hypoxia
P5	36	F	16	0.5	Severe	ABI: Hypoxia
P6	58	F	10	2	Severe	ABI: Hypoxia
P7	46	M	14	2	Severe	ABI: Hypoxia
P8	34	F	13	1	Severe	ABI: Hypoxia
P9	19	M	12	1	Severe	ABI: RTA
P10	57	F	5	11	Severe	ABI: Hypoxia
P11	20	M	13	0.6	Severe	TBI: RTA
P12	24	M	13	2	Severe	TBI: RTA

Abbreviations: TPO time post onset; RTA road traffic accident; ICH intracranial haemorrhage

Participants took part in a feasibility study to determine whether a peer-mediated training intervention improves social communication outcomes. They were randomly assigned to either a peer-led intervention group or a staff-led activity group. Both groups met twice a week for 8 weeks. Ethical approval for the study was gained from South Central Hampshire ‘A’ NHS Research Ethics Committee (reference 14/SC/0048). All participants (or their representatives) gave informed, written consent to participate in this investigation.

For both intervention conditions, group meetings were filmed according to a pre-determined protocol using four tripod mounted GoPro Hero 3 edition camcorders in order to capture interactions from multiple angles. An EDITIGE ETM-001 dual

microphone was attached to each camera. Ten minute clips of participants engaging in group interaction were prepared for analysis using Final Cut Pro editing software version 10.2.3 (Apple Inc). Each clip presented three-way views of a participant interacting with the rest of the group. Clips were labelled using pseudonyms and copied onto an encrypted hard drive in random sequence to conceal the time point of data collection. Thirty-one clips were rated using the four outcome measures. The measurement order was also randomised.

Measurement scoring procedures

The PPIC (Linscott et al.,1996) assesses 84 social communication behaviours organised across 10 sub-scales. Specific behaviour frequencies are scored on a 4-point rating scale, where zero indicates 'not at all' and three indicates 'almost always/always'. The scoring system also includes a 6-point feature summary scale (FSS) (0-5), where zero is normal and 5 very severely impaired. Inter-rater reliability findings have previously been reported on feature summary scale data only, and this procedure was followed in this investigation.

Social communication behaviours on the BRISS-R (Farrell, Rabinowitz, Wallander and Curran, 1985) are rated on a 7-point Likert Scale, where 1 indicates very inappropriate behaviours and 7 indicates fully appropriate behaviours. Behavioural descriptors are included to aid scoring (e.g. on the PCSS use of humour sub-scale, 'childish/excessive humour some of the time' scores 2; on the PDBS self-centred behaviour sub-scale, 'talked about self some of the time' scores 5).

The MPC comprises two sub-scales, interaction and transaction, both tailored to reflect the social communication needs of people with TBI. On both sub-scales raters

score a participant on a 9-point Likert scale (0-4 with a half point scoring option, and where zero indicates no participation). Questions guide the rater (e.g. ‘Does the person add information to maintain the topic?’; ‘Do they present information in an organised way?’). Descriptors (or anchors) accompany the scale (e.g. on the interaction sub-scale, ‘no attempt to engage with the communication partner’ scores zero; on the transaction sub-scale, ‘consistently conveys content to achieve the task purpose’ scores 3).

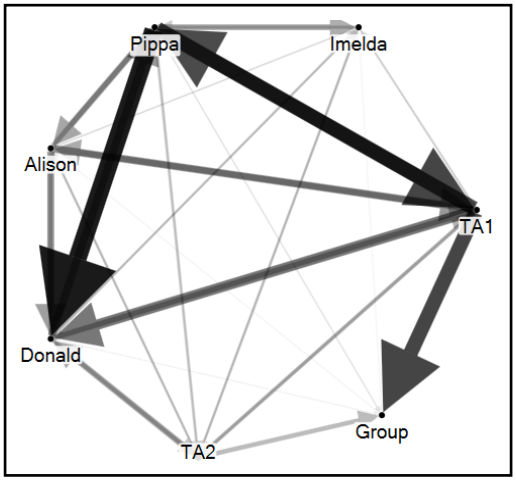
In contrast with the PPIC, BRISS-R and MPC, which are established observational rating scales designed to evaluate dyadic interactions, the INT is a digital measure designed to measure communication behaviour frequencies. Interaction patterns are captured using a 14-item behaviour coding system (see Table 2) comprised of verbal and non-verbal initiations and responses.

Table 2: INT behaviour codes

1	Initiation to one other - verbal
2	Initiation to group - verbal
3	Initiation: eye gaze (people) - non-verbal
4	Initiation: eye gaze (objects) - non-verbal
5	Initiation – non-verbal: pointing/reaching/gesturing
6	Initiation - non-verbal: facial expression
7	Response - verbal (1 word) to one other
8	Response - verbal (1 word) to group
9	Response - verbal (more than 1 word) to one other
10	Response - verbal (more than 1 word) to group
11	Response – non-verbal: head nod/shake
12	Response – non-verbal: pointing/reaching/gesturing
13	Response – non-verbal: facial expression
14	Other vocal response: laughter, scream, singing, fillers (e.g.um)

For example, interaction code 13 (response non-verbal: facial expression) is intended to capture all facial expressions (voluntary or involuntary) used in response to another participant; interaction code 2 (initiation to group – verbal) is intended to capture all verbal initiations to the group, but does not detail the particular utterance type (such as a question or statement). Network visualisations can be generated to show who interacts with who, whether particular individuals dominate the conversation, and identify initiators (who draw others into the conversation) and responders (see Figure 1). This is a new group interaction measure and this is the first report on its reliability as a group interaction measure.

Figure 1: An INT visualisation showing the pattern of connections between group interactors.



Arrows connecting participants indicate conversation direction. Lines connecting participants become thicker and more colour dense as initiation and response frequencies increase 'Group' is also included as a destination node to distinguish between paired connections and connections between an individual and the group. TA= therapy assistant.

Figure 1 shows a group comprising four participants with ABI, supported by two therapy assistants. Pippa is a dominant speaker, but talks mostly to Donald and TA1. The participants engage individually, and the frequency of participant initiations to the group as a whole is low. TA1 initiations to the group as a whole are high.

Rating procedure

Four raters were recruited and trained to independently rate interactional data using the four outcome measures. Two raters (masters-level speech and language therapy students at UCL) evaluated 24 films. Two further raters (a masters-level graduate in Human Communication Science and a recently qualified speech and language therapist) were recruited and trained to evaluate the remaining seven films. All raters were familiar with the profile of impairment seen in ABI, but unfamiliar with the outcome measures used for this study. Raters received approximately 5 hours of

training over two sessions to familiarise them with the outcome measures. The raters were trained in pairs, and the training was the same for both pairs. The format of the training was the same for all measures. Raters were trained to use two measures in each training session. The training was designed to familiarise the raters with the measures and provide practice using the scales to rate video footage of an ABI group interaction that was not associated with this investigation. Rater performance was not measured against criteria during the training. Practice tasks were structured, and discrepancies were discussed after each practice task in the group training sessions. Independent practice tasks were completed between sessions using new group video footage to consolidate the training. On independent tasks, raters conferred between themselves to address any inconsistencies and contacted the project team to discuss unresolved discrepancies. Following training, the rater pairs independently evaluated the sampled data using the four outcome measures. The measurement order was randomised.

Data analysis procedure

The four measures comprised a total of 32 scales. In this analysis of pilot data, scales were aggregated to reduce the number of comparisons. This aggregation procedure reduced the total number of rating scales to seven. Following the procedure documented by Braden et al. (2010), the 10 subscales of the PPIC were summed for each participant to provide an overall summary score from FSS data. For the BRISS-R, data from the three subscales of the PCSS and PDBS were summed for each participant to derive one PCSS and one PDBS summary score. For the INT, scores for each participant were summed to provide an overall score for initiations and an overall score for responses.

Inter-rater reliability was calculated using intra-class correlations (ICC) (Shrout and Fleiss, 1979). ICCs calculate the degree of variance between raters in relation to the variation in scores. A high level of agreement therefore requires low rater variation on a wide variation in scores. Rater reliability was measured using ICC type 3,1. This type is defined in SPSS as two-way mixed. The mixed effects model was chosen because the participants were randomly selected but the raters were fixed. 'Single measures' and 'absolute agreement' were selected in order to determine the generalisable accuracy and reliability of scores.

ICC values generally fall between 0 and 1, with 1 defined as perfect agreement, 0 defined as random agreement and minus numbers indicating disagreement (Hallgren, 2012). Guidelines for interpretation vary. Previous investigators (e.g. McDonald, Tate et al., 2008; Togher et al., 2010) report reliability using the values proposed by Cicchetti (1994). According to this guideline, values below 0.4 are considered poor, values between 0.4 and 0.59 are considered fair, values between 0.6 and 0.74 are considered good, and excellent for values at 0.75 and above. However Post (2016), in an editorial comment, has questioned the lenience of the existing guidelines for interpreting ICCs, and Koo and Li (2016) have proposed more stringent criteria for magnitude of agreement. According to these new interpretation guidelines, values below 0.5 indicate poor reliability, between 0.5 and 0.75 are considered moderate, between 0.75 and 0.9 are considered good, and values above 0.9 indicate excellent reliability. Further, as reliability studies report an ICC estimate, Koo and Li (2016) recommend using confidence intervals to report probability coverage rather than the point estimate alone. These recommendations for

interpretation using more stringent criteria, and to report both ICC estimate and confidence intervals have been followed in this report.

Results

Results for intra-class correlation calculations for all primary outcome measure sub-scales are presented in Table 3. 95% confidence intervals are reported.

Table 3: Intra-class correlations (ICC 3,1) with confidence intervals (CI) 95% for the outcome measures

Outcome Measure	ICC (3,1)	95% CI
MPC		
Interaction	0.77	0.58 – 0.88
Transaction	0.74	0.52 – 0.87
BRISS-R		
Personal conversational style (PCSS)	0.49	0.32 - 0.63
Partner directed behaviour (PDBS)	0.50	0.33 - 0.64
PPIC	0.68	0.61 – 0.73
INT		
Initiations	0.83	0.68 – 0.91
Responses	0.69	0.45 – 0.84

.

Abbreviations: MPC Adapted Measure of Participation in Conversation; BRISS-R Behaviourally Referenced Rating System of Intermediate Social Skills; PCSS Personal Conversational Style Scale; PDBS Person Directed Behaviour Scale; PPIC Profile of Pragmatic Impairment in Communication; INT Interactional Network Tool

Using the criteria from Koo and Li (2016), the ICC estimate for the MPC interaction sale (ICC=0.77) and transaction scale (ICC=0.74) was moderate to good. The ICC estimate for the INT initiations (ICC=0.83) was moderate to excellent and for the INT responses (ICC=0.69) was poor to good.

ICC estimates of reliability for the BRISS-R PCSS scale (ICC=0.49) and PDBS scale (ICC=0.50) were poor to moderate. The estimate for the PPIC scale (ICC=0.68) was moderate. However, the narrow confidence interval (CI=0.61-0.73) on the PPIC was atypical in comparison to the CI ranges for the other outcome measures suggesting lower variability of score distribution. On inspection, the data showed presence of skew. Forty-two percent of the scores fell into the category of zero (median score=1; range 0-5). These findings potentially undermine the validity of the application of the PPIC and BRISS-R to group interaction data.

Inter-rater reliability was adequate for the MPC interaction and transaction scales and the INT initiations. Findings for the INT responses were less clear given the wide confidence interval (ICC=0.69; 95% CI=0.45-0.84). Although inconclusive, this is a new measure and data from future studies is required to evaluate consistency of findings. Both the MPC and the INT achieved these outcomes without extensive training and both were selected for the subsequent trial of a peer-led intervention for social communication skills. The BRISS-R and the PPIC were not selected. The BRISS-R scales indicated only a poor to moderate agreement. It has been acknowledged by previous researchers that a high level of rater agreement is not easily achieved with this measure, even in dyadic coding and with extensive rater training (Togher, McDonald, *et al.*, 2014). The PPIC showed a moderate level of

agreement, but scoring anomalies in the application to group encounters reduced the range of variance in the ratings and undermined reliability with group data.

Discussion

This evaluation showed differential inter-rater reliability across the BRISS-R, PPIC, MPC and the INT. Findings for the BRISS-R and the PPIC indicate insensitivity to the group interaction construct. This is unsurprising as both were designed to evaluate isolated and participant-specific behaviours. ICCs for the BRISS-R indicated only poor to moderate agreement. ICC findings on the PPIC were undermined by the reduced range of variance, suggesting a mismatch between the measure and the group interaction construct under investigation. Scoring anomalies on items in both measures also reflected contrasts between dyadic and group conversation norms. For example, on the BRISS-R social manners scale, multiple interruptions are rated as inappropriate. These occurred in the group conversations as a means to negotiate a turn. On the literal content subscale of the PPIC, ‘leaves out parts of sentences’ were often a natural result of conversational cross-talk in a group rather than an unintended outcome from a poorly constructed sentence. Anomalies such as these undermined rater agreement.

Although the procedure to aggregate scores enabled a reduction in the number of comparisons, it also limited our understanding of variation between each component of the subscale. A future factor analysis would enable us to more formally investigate the variance within each sub-scale component, providing a clear picture of the contribution of each variable to the outcome.

The MPC is a global measure of participation in a conversational exchange and therefore contrasts with other dyadic interaction measurement tools that evaluate change in discrete behaviours. ICCs for the MPC interaction and transaction scales applied to the group interaction data were moderate to good. This outcome contrasts with findings from Togher et al. (2010) for dyadic interactions where both scales achieved excellent agreement, although the more lenient interpretation guidelines of Cicchetti (1994) were applied in that investigation. Togher et al. suggest that the strength of rater agreement in their study may reflect the ease of rating structured and predictable dyadic conversations with familiar people. The group conversations evaluated in this investigation involved a mix of staff-led interactions consisting of guided discussion (with the predictability of an implicit structure), and peer-led group interactions comprising independent discussion without staff present. Findings indicate an acceptable level of agreement rating both discourse types, although further research across a range of discourse types and contexts is required.

The INT provides a new way to measure change in social participation. It uses relational data in the form of communication behaviour frequencies coded by raters. It is a measure of group participation, and the frequency counts show patterns of interaction between participants revealing how they communicate (and cue each other to communicate) rather than what they communicate. As such, the strength of the ties between participants may be counts based on positive or negative expressions of agreement or disagreement. Rater feedback indicated that the interaction types that make up the coding system were representative of the group conversation behaviours observed in the filmed interactions. The INT does not

examine conversation content or index the quality of the individual behaviours. Such an evaluation would require an established qualitative methodology.

The INT provides a primarily quantitative analysis. However, in the context of clinical work, it provides a qualitative perspective on group dynamics. The behavioural frequency data provide a detailed quantitative measure of the relational contacts between participants. For clinicians the network graphs provide a qualitative interpretation, enabling new insights into adaptive behaviour and salient feedback for participants (e.g. as a means to set goals to be less dominant or more inclusive). There is application for trainee clinicians to address the potentially negative effect of professional power on interactions, and for group facilitators to devise action plans (e.g. to draw specific individuals into discussion or to pair individuals with stronger ties to those with weaker ties). At the same time, the visualisations (pre-post intervention) can be used to provide quantifiable evidence of change in review reports for funding bodies or discharge reports for new clinical teams.

Analyses for this investigation were based on a pooling of behavioural frequency counts for each participant across each measurement stage and inter-rater reliability for initiations was moderate to excellent and responses was poor to good. Additional work is needed to evaluate rater consistency across individual codes, providing further information on the reliability and precision of the tool. Although inter-rater reliability is an important indicator of consensus across raters, further research is required to validate use of individual codes across different communication environments and interaction types (involving peers and/or neurotypical conversation partners).

A potential limitation of observational ratings or frequency counts as a method of conversation outcome measurement is that their presence or absence in a given conversation sample may be dependent on the length of video sample. Inter-rater reliability in this evaluation was calculated on 10 minute samples. In previous social communication skills investigations in ABI, sample lengths have ranged from 5 minutes (Togher et al., 2013) to 15 minutes (Helffenstein and Wechsler, 1982). However, these were dyadic conversations, and the ability of a 10 minute sample to capture a representative snapshot of conversation between four and eight individuals with variable interpersonal communication skills cannot be guaranteed. Future research to investigate the effects of peer group conversation sample lengths on measurement outcomes is therefore warranted to establish optimal sample duration.

In summary, this investigation examined the inter-rater reliability of three existing conversation measures, previously used with dyadic data, to assess group interaction data. The INT, a new digital measure, developed specifically for group settings, was also evaluated. Findings show that the MPC and the INT may be potentially useful tools to measure participation in groups. The INT offers a new approach to analysing patterns of social interaction and measuring outcomes in group settings, with a novel focus on capturing social connections between people. Further development and validation of this instrument is indicated.

Acknowledgements

The authors thank Adam Searle of TGC Consulting Ltd for his support with the INT, and Anthony Geffen and the team at Atlantic Productions for their advice on the filming protocol.

Susan Howell is supported by a postdoctoral award from the Economic and Social Research Council [grant reference ES/T008504/1].

Disclosure statement

The authors report no conflicts of interest.

References

1. Bandura, A. (1971). *Social learning theory*. New York: General Learning Press.
2. Braden, C., Hawley, L., Newman, J., Morey, C., Gerber, D. and Harrison-Felix, C. (2010). Social communication skills group treatment: a feasibility study for persons with traumatic brain injury and comorbid conditions. *Brain Injury*, 24(11), pp. 1298–1310. doi: 10.3109/02699052.2010.506859.
3. Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), pp. 284–290. doi: 10.1037/1040-3590.6.4.284.
4. Cicerone, K. D. and Azulay, J. (2007). Perceived self-efficacy and life satisfaction after traumatic brain injury. *Journal of Head Trauma Rehabilitation*, 22(5), pp. 257–266. doi: 10.1097/01.HTR.0000290970.56130.81.
5. Coelho, C. A., Liles, B. Z. and Duffy, R. J. (1991). Analysis of conversational discourse in head-injured adults. *Journal of Head Trauma Rehabilitation*, 6(2), pp. 92–98. doi: 10.1097/00001199-199106000-00011.
6. Coelho, C. A., Youse, K. M. and Le, K. N. (2002). Conversational discourse in closed-head-injured and non-brain-injured adults. *Aphasiology*, 16(4–6), pp. 659–672. doi: 10.1080/02687030244000275.
7. Coulthard, M. (1984). Conversation analysis and social skills training. In Trower, P. (ed.), *Radical Approaches to Social Skills Training*. The University of Michigan: Croom Helm.
8. Dahlberg, C. A., Cusick, C .P., Hawley, L.A., Newman, J.K., Morey, C.E.,

Harrison-Felix, C.L. and Whiteneck, G.G. (2007). Treatment efficacy of social communication skills training after traumatic brain injury: a randomized treatment and deferred treatment controlled trial. *Archives of Physical Medicine and Rehabilitation*, 88(12), 1561–1573. doi: 10.1016/j.apmr.2007.07.033.

9. Douglas, J. M., Knox, L., De Maio, C., Bridge, H., Drummond, M.,Whiteoak, J. (2019). Effectiveness of Communication-specific Coping Intervention for adults with traumatic brain injury: preliminary results. *Neuropsychological Rehabiliation*, 29(1), pp. 73-91, DOI I 10.1080/09602011.2016.1259114
10. Eggins, S. and Slade, D. (2005). *Analysing Casual Conversation.* London, Equinox.
11. Farrell, A., Rabinowitz, J. A., Wallander, J. L. and Curran, J.P. (1985). An evaluation of two formats for the intermediate-level assessment of social skills. *Behavioral Assessment*, 7, pp. 155–171.
12. Finch, E., Copley, A., Cornwell, P.and Kelly C. (2016). Systematic review of behavioral interventions targeting social communication difficulties after traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 97(8), pp. 1352–1365. doi: 10.1016/j.apmr.2015.11.005.
13. Finch, E., Cornwell, P, Copley, A., Doig, E. and Fleming, J. (2017). Remediation of social communication impairments following traumatic brain injury using metacognitive strategy intervention: a pilot study. *Brain Injury*, 31(13–14), pp. 1830–1839. doi: 10.1080/02699052.2017.1346284.
14. Flanagan, S., McDonald, S. and Togher, L. (1995). Evaluating social skills following traumatic brain injury: the BRISS as a clinical tool. *Brain Injury*, 9(4), pp. 321–338. doi: 10.3109/02699059509005773.

15. Grice, P. (1975). Logic and conversation. In: Cole, P. and Morgan, J. (eds.). *Studies in Syntax and Semantics,* Vol.3. New York: Academic Press, pp. 183–198.
16. Hallgren, K.A. (2012). Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), pp. 23–34. doi: 10.20982/tqmp.08.1.p023.
17. Hammond, F., Barrett, R., Dijkers, M.P., Zanca, J.M., Horn, S.D., Smout, R.J., Guerrier, T., Hauser, E. and Dunning, M.R. (2015). Group therapy use and its impact on the outomes of inpatient rehabilitation following traumatic brain injury. *Archives of Physical Medicine and Rehabilitation*, 96(80), pp. S282-292. doi: 10.3109/10641955.2015.1046604.
18. Hartley, L. L. (1995). *Cognitive-Communicative Abilities Following Brain Injury*. New York: Delmar, Cengage Learning.
19. Helffenstein, D. A. and Wechsler, F. S. (1982). The use of interpersonal process recall (IPR) in the remediation of interpersonal and communication skill deficits in the newly brain-injured. *Clinical Neuropsychology*, 4, pp. 139–142.
20. Howell, S. (2018). Measuring outcomes from a peer-led social communication skills intervention for adults following acquired brain injury. Thesis (PhD), UCL (University College London). https://discovery.ucl.ac.uk/id/eprint/10059713
21. Howell, S., Beeke, S., Pring, T. and Varley, R. (2020). Measuring outcomes of a peer-led social communication skills intervention for adults with acquired brain injury: A pilot investigation. Neuropsychological Rehabilitation. doi: 10.1080/09602011.176892

22. Koo, T. K. and Li, M.Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), pp. 155–163. doi: 10.1016/j.jcm.2016.02.012.
23. Lave, J. and Wenger, E. (1991). *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press. doi: 10.2307/2804509
24. Linscott, R.J., Knight, R.G. and Godfrey, H.P.D. (1996). The Profile of Functional Impairment in Communication (PFIC): a measure of comunication impairment for clinical use. *Brain Injury*, 10 (6), pp. 397- 412.
25. MacDonald, S. (2017). Introducing the model of cognitive-communication competence: A model to guide evidence-based communication interventions after brain injury. Brain Injury, 31 (13-14): 1760-1780, doi: 10.1080/02699052.2017.1379613
26. McDonald, S., Tate, R., Togher, L., Bornhofen, C., Long, E., Gertler, P. and Bowen, R. (2008). Social skills treatment for people with severe, chronic acquired brain injuries: a multicenter trial. *Archives of Physical Medicine and Rehabilitation*, 89(9), pp. 1648–1659. doi: 10.1016/j.apmr.2008.02.029.
27. Post, M.W. (2016). What to do with “moderate” reliability and validity coefficients? *Archives of Physical Medicine and Rehabilitation*, 97(7), pp.1051-1052. doi.org/10.1016/j.apmr.2016.04.001
28. Scott, J. (2017). *Social Network Analysis*. 4th ed. London: SAGE Publications Ltd.
29. Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, 86(2), pp. 420–428. doi: 10.1037/0033-2909.86.2.420

30. Sohlberg, M.M., MacDonald, S., Byom, L., Iwashita, H., Lemoncello, R., Meulenbroek, P., Ness, B. and O'Neil-Pirozzi, T.M. (2019). Social communication following traumatic brain injury part I: State-of-the-art review of assessment tools, *International Journal of Speech-Language Pathology*, 21:2, 115-127, DOI: 10.1080/17549507.2019.1583280
31. Tate, R.L. (2019). Measuring outcomes and monitoring progress in the era of evidence-based clinical practice. Brain Impairment, 20, pp.276-288. doi:10.1017/BrImp.2019.28
32. Teasdale, G. and Jennett, B. (1974). Assessment of coma and impaired consciousness. A practical scale. *The Lancet*, 304(7872), pp. 81–84. doi: 10.1016/S0140-6736(74)91639-0.
33. Togher, L., McDonald, S., Tate, R., Power, E. and Rietdijk, R. (2013). Training communication partners of people with severe traumatic brain injury improves everyday conversations: a multicenter single blind clinical trial. *Journal of Rehabilitation Medicine*, 45(7), pp. 637–645. doi: 10.2340/16501977-1173.
34. Togher, L., Power, E., Tate, R., McDonald, S. and Rietdijk, R. (2010). Measuring the social interactions of people with traumatic brain injury and their communication partners: the adapted Kagan scales. *Aphasiology*, 24 (6-8), pp. 914–927. doi: 10.1080/02687030903422478.
35. Togher, L., Wiseman-Hakes, C., Douglas, J., Stergiou-Kita, M., Ponsford, J., Teasell, R., Bayley, M. and Turkstra, L.S. (2014). INCOG recommendations for management of cognition following traumatic brain injury, part IV: cognitive communication. *Journal of Head Trauma Rehabilitation*, 29(4), pp. 353–368. doi: 10.1097/HTR.0000000000000071.

36. Trower, P., Bryant, B. and Argyle, M. (1978). *Social Skills and Mental Health*. Hove, UK: Routledge.
37. Vygotsky, L.S. (1978). *Mind in Society: the Development of Higher Psychological Processes.* Cambridge, Massachusetts: Harvard University Press.
38. Wasserman, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications.* Cambridge: Cambridge University Press.