Neuropsychobiology

Research Article

Neuropsychobiology
DOI: 10.1159/000523778

Received: May 20, 2021
Accepted: February 11, 2022
Published online: April 1, 2022

Validation of the Apraxia Screen TULIA (AST) in Schizophrenia

Hanta Bachofner[a] Konstantin A. Scherer[a] Tim Vanbellingen[b]
Stephan Bohlhalter[b, c] Katharina Stegmayer[a] Sebastian Walther[a]

[a]Translational Research Center, University Hospital of Psychiatry, University of Bern, Bern, Switzerland; [b]Neurocenter, Luzerner Kantonsspital, Luzern, Switzerland; [c]University Hospital, University of Zurich, Zurich, Switzerland

Keywords
Schizophrenia spectrum disorders · Hand gestures · TULIA · AST · Nonverbal communication

Abstract
Introduction: Deficits in social interaction and community functioning, including impaired use, performance, and perception of hand gestures, are key features in schizophrenia. A well-established tool to assess gesture deficits is the test of upper limb apraxia (TULIA). However, given its time-consuming application based on video analyses, research has proposed the bedside apraxia screen of TULIA (AST). This study aims to test the validity and reliability of the AST to detect gesture abnormalities at bedside in a sample of 27 patients diagnosed with schizophrenia, schizotypal disorder, acute and transient psychotic disorders, or schizoaffective disorder. ***Methods:*** Patients completed the 48-item TULIA and the 12-item AST. Two different raters assessed the AST: one at bedside (online) and the other based on the video recordings. ***Results:*** The total AST scores demonstrated a high parallel reliability, moderate inter-rater reliability on a single-item level, and good construct validities. ***Conclusions:*** The psychometric properties of the AST suggest it can well be used for the clinical assessment of gesture deficits in schizophrenia. However, when detailed information is required, the AST rated from video or conducting the full TULIA is recommended. The findings call for refining the selection of the TULIA items for a psychosis-AST bedside test to increase specificity.

Introduction

Ranked among the top 20 causes of disability worldwide, schizophrenia is a highly disabling disorder [1]. Although the etiopathology varies among individuals, schizophrenia is typically persistent. One of the most prominent features is impaired social cognition, i.e., difficulties identifying the emotions or intentions of others and responding emotionally to others [2]. Importantly, impaired social cognition and negative symptoms are predictors of poor functional outcomes in schizophrenia [3, 4]. Impaired social role functioning is characterized by difficulties in fulfilling basic social roles such as being a spouse, parent, or employee. Up to two-thirds of people with schizophrenia are unable to establish or maintain

Karger@karger.com
www.karger.com/nps

Correspondence to:
Hanta Bachofner, hanta.bachofner@upd.unibe.ch

these roles even when remitted from psychotic symptoms [5]. People with schizophrenia spectrum disorders show remarkably stable long-term impairments in social functioning – with differences among trajectories already evident in childhood [6]. In terms of the disorder and its treatment, social functioning has been recognized as a key outcome marker [7].

Patients with schizophrenia show deficits in perceiving relevant cues in nonverbal communication [8]. Examples are impairments in the perception of incidental movements as gestures or the false classification of neutral gestures as a threat [9, 10]. Likewise, there are disturbances in imitation tasks, such as replicating meaningless manual and oral gestures or imitating affective facial expressions [11].

The correct use of hand gestures may indicate the level of social functioning [12]. Hand gestures, which are key to nonverbal communication, substitute or support verbal information [13] and rely both on motor and language skills [8, 14, 15]. Accumulating evidence demonstrates gesture impairments, including hand gesture impairments throughout all stages of schizophrenia, from the prodrome to the chronic state [16–22]. Specifically, subjects at risk for psychosis tend to use co-speech gestures incongruent to the speech content, rendering the information ambiguous [21]. Generalized nonverbal deficits have been detected in patients with schizophrenia spectrum disorders. Besides impaired interpretation, schizophrenia is associated with compromised production of gestures, which arise from generalized sensory and motor deficits [22]. Both these functional domains, production and perception, are associated with motor abnormalities, frontal lobe dysfunction, impaired working memory, as well as positive symptoms in schizophrenia [20, 22–24]. Even when researchers controlled for nonverbal intelligence and working memory deficits, performance in gesture tasks remained strongly impaired in schizophrenia [22]. Furthermore, gesture deficits are also linked to negative symptoms [18, 25] and have been found to predict an unfavorable course over 6 months with stable negative symptoms and a decline in social functioning [26]. In sum, the assessment of gesture functioning may provide critical information on the expected outcome of schizophrenia.

Spontaneous gesturing is distinct from gesturing on command. Gestures may be tested in two principal domains: imitation (repeat a gesture demonstrated by the examiner) and pantomime (production following the verbal command of the examiner). Gesture categories include nonsymbolic, meaningless (e.g., "put index finger on top of nose"), or symbolic, meaningful gestures. Such meaningful gestures are divided into intransitive, communicative (e.g., "salute like a soldier" or "point to a bird in the sky"), or transitive, object-related (e.g., "use a hammer" or "comb your hair") [27]. The test of upper limb apraxia (TULIA) is a valid tool to assess gesture performance [27]. Using the TULIA, gesture deficits in the pantomime domain have been found in up to 66% of the patients with schizophrenia and in up to 33% in the imitation domain [19, 20, 22, 24].

Despite TULIA's ability to identify gesture deficits in schizophrenia, it is currently not implemented in clinical practice because the rating from a video is time consuming and requires training as each item may receive a score of 0–5. Thus, a short yet valid bedside test is preferable. Research in apraxia has demonstrated that the adapted short version, i.e., apraxia screen of TULIA (AST, Vanbellingen et al. [28]), is less time consuming, and physicians may evaluate gesture performance while testing ("online," "at bedside") without using a camera due to a simplified dichotomous scoring system. The AST could be an important screening tool, addressing the detection of the onset of psychosis, given that subjects at clinical high risk for psychosis and patients with first episode psychosis already demonstrate critical gesture deficits [21, 29].

This study aims to analyze whether the AST can be as efficiently used as the full TULIA in screening for gesture deficits in schizophrenia. Furthermore, we wanted to explore whether AST bedside and video ratings correlated and hypothesized strong correlations in schizophrenia patients. Given the AST ratings prove to correspond to those of the TULIA, it could be used in research protocols or clinical evaluations as a novel, more economical tool to identify gesture impairments.

Materials and Methods

Participants

Patients were recruited from the in- and outpatient departments of the University Hospital of Psychiatry Bern, Switzerland. In total, the study comprised 29 German-speaking patients between 18 and 65 years of age who had been diagnosed with schizophrenia spectrum disorders according to ICD-10 criteria: schizophrenia (F20), schizotypal disorder (F21), acute and transient psychotic disorders (F23), or schizoaffective disorder (F25). Two patients completed the bedside test AST but refused to complete the full TULIA. Therefore, they were excluded, leaving a total of 27 participants for further analysis (see Table 1). All but four participants were outpatients at the time of testing. Exclusion criteria were substance dependence or any medical/neurological condition interfering with gesture performance, e.g., stroke. According

Table 1. Characteristics of patients

Age, years	46.6 (SD = 12.9)*
Gender	
Women	11
Men	16
Diagnoses	
Schizophrenia (F20)	19
Schizotypal disorder (F21)	2
Acute and transient psychotic disorders (F23)	2
Schizoaffective disorder (F25)	5
Episodes, *n*	5.9 (±6.1)*
Duration, years	19.7 (±12.9)*
PANSS	
total	80.3 (±21.3)*
positive scale	18.3 (±6.6)*
negative scale	21.2 (±7.2)*
general scale	40.7 (±10.3)*
Drugs	
CPZ, mg	432.4 (±337.2)*
Mood stabilizers, *n*	6
Antidepressants, *n*	8
Z-Drugs/benzodiazepines, *n*	2
Illegal drugs (recreational use of cannabis or cocaine)	
total, *N*	4
Total scores	
AST bedside_dichotomous	7.5 (±2.3)*
AST video_dichotomous	9.0 (±3.0)*
AST video	45.8 (±11.7)*
TULIA	199.4 (±33.6)*

PANSS, Positive and Negative Syndrome Scale; CPZ, chlorpromazine. * Mean (standard deviation [SD]).

to the Edinburgh Handedness Inventory [30], 24 patients were right- and three were left-handed. At the time of the study, 2 patients were off antipsychotic medication. For all other participants, chlorpromazine (CPZ)-equivalent doses of their medication were calculated according to Woods (2003) [31]. All patients provided written informed consent. The protocol adhered to the Declaration of Helsinki, 1975, had been approved by the local Ethics Committee.

Procedures

The TULIA includes 48 items in the domains of imitation and pantomime. Each domain can be subdivided into three categories: meaningless, intransitive (symbolic), and transitive (tool-related) gestures. The strong correlation ($r = 0.82$) between the TULIA and the apraxia test by De Renzi et al. [32] indicates that the two scales measure related constructs of gesture production. The TULIA was validated with the apraxia test by De Renzi et al. [32] and proved mostly good to excellent results both at the level of the six subtests as well as at the individual item level, in terms of high internal consistency, inter-, and intrarater reliability [27]. Total TULIA scores range between 0 and 240, with lower scores indicating poor performance. The TULIA was originally developed and validated for stroke patients. For these patients, the threshold for apraxia was set at a total score of 194 points, two standard deviations below the mean score of 217.5 of controls (healthy adults with a mean age of 61 years). Accordingly, cut-off levels for moderate (<130 points) and for severe apraxia (<60 points) were defined. For administration in patients with schizophrenia, Walther et al. [19] suggested an adjusted TULIA cut-off score of 210 points for gesture deficits when they tested TULIA performance in comparison to an age- and gender-matched control group with a mean age of 40 years. Without matching for age, TULIA results are biased, because elderly people generally score lower than younger ones [19]. With the adjusted (i.e., higher) cut-off score, two out of three (66.7%) schizophrenia patients showed impairments, affecting all categories of the TULIA [19, 22], compared to 40% with the original cut-off that was calculated in a study of apraxia patients [17, 20, 22].

Instruction and execution of the TULIA require approximately 20 min [27], and the rating thereof requires approximately 25 min. TULIA performance is rated item-by-item from the video recordings using a 6-point scoring method (range 0–5) as indicated in the manual [27]. In the lower range (0–2), content errors (e.g., substitutions and perseverations) and temporal and spatial errors (e.g., body-part-as-object errors, errors in spatial orientation, omissions, overshoot, and extra movements) are considered. The higher score range (3–4) includes minor temporal and spatial errors – corrected by the patient or not – as well as slight changes in movement trajectory. The highest score of 5 is given for entirely correct gestures.

Bedside evaluation was not part of the concept when TULIA was created. However, an experienced rater may score the TULIA instantaneously when the performance is slow. To foster time-efficient assessment in clinical settings, the AST was introduced, which takes approximately 3 min to administer [28].

The 12 items of the AST were selected by an item-reduction analysis from all TULIA items. In the AST, items 1–7 belong to the imitation domain and items 8–12 to the pantomime domain. In these domains, the AST – as the full TULIA – includes a nonsymbolic, meaningless gesture (AST-item 1), intransitive, communicative gestures (AST-items 2, 8, and 9), and transitive, tool-related gestures (AST-items 3–7 and 10–12) [28]. The validation of the 12-item AST with the full, 48-item TULIA showed remarkable diagnostic reliability with high specificity (100%), sensitivity (95%), positive predictive value (100%), and negative predictive value (92%) for the presence and severity of apraxia in patients with a history of stroke. Furthermore, scores of the AST and the full TULIA correlated strongly (Pearson correlation coefficient $r = 0.96$; $p < 0.001$), as did the AST scores with their corresponding 12 items of the TULIA (test-retest reliability of $r = 0.95$, $p < 0.001$). Thus, the AST has proven to be reliable, valid, and cost-effective in post-stroke apraxia, which can be conducted within several minutes [28]. The total AST score ranges from 0 to 12, with low scores indicating a gesture deficit and high scores indicating a better accomplishment. In comparison, the cut-off levels are 5 for severe and 9 for a mild version of apraxia in patients with brain damage following a stroke [28].

The participants performed the gestural tests TULIA and AST with their dominant hand. For both gestural tests, they were seated face-to-face with the examiner and rested their forearms on the table positioned between them. Participants were instructed to imitate gestures demonstrated by the examiner or to pantomime gestures following verbal commands. Special care was taken to ascertain that the participant understood every verbal instruction. First,

Table 2. Sensitivity, specificity, and predictive values of the AST bedside and the full TULIA

Gesture deficit	TULIA		Total	
	no	yes		
AST bedside				
No	7	3	10	Negative predictive value = 70.0%
Yes	7	10	17	Positive predictive value = 58.8%
Total	14	13	27	
	Specificity = 50.0%	Sensitivity = 76.9%		

Table 3. Sensitivity, specificity, and predictive values of the AST video and the full TULIA

Gesture deficit	TULIA		Total	
	no	yes		
AST video				
No	9	2	11	Negative predictive value = 81.8%
Yes	5	10	15	Positive predictive value = 66.7%
Total	14	12	26	
	Specificity = 64.3%	Sensitivity = 83.3%		

the structured AST and then the TULIA (for detailed procedure see Vanbellingen et al. [28]; Vanbellingen et al. [27]) were conducted and recorded on video. The AST was rated immediately (bedside, by K.S.) during the examination, while the TULIA was rated afterward from the video recordings. While conducting the AST, the examiner rated the participant's performance in a dichotomous score (pass = 1 vs. fail = 0). In fact, passing an AST item corresponds to a TULIA item rating of 3–5, while failing on the AST corresponds to a TULIA item rating of 0–2.

In this study, the video recordings of the AST and the TULIA were rated by an independent expert, entirely blinded to patient status (H.B.). She had never met the patients or had any information about their condition or medical history, nor did she have access to the results of the AST bedside ratings. The construct validity and the test-retest reliability between the AST and the TULIA were assessed based on independent ratings (H.B. and S.W.).

After the AST and TULIA assessments, the Positive and Negative Syndrome Scale (Kay, Fiszbein and Opler, 1987) was conducted to assess general symptom severity. Both, the person conducting the AST and the TULIA examinations (K.S.) as well as the person who scored the TULIA and the AST from video (H.B.), were trained by the senior author (S.W.) to achieve optimal inter-rater reliability.

Statistical Analysis

First, parallel reliability between total scores of the bedside-rated AST versus the video-rated AST was evaluated with Spearman's rho correlation. By calculating Cohen's kappa for the AST rating performed at bedside versus from video, we tested inter-rater reliability at the individual item level. Sensitivity and specificity between bedside and video AST ratings (dichotomous in each case) at the single-item level were calculated using crosstabs in SPSS. For construct validity, Spearman's rho correlation analysis was calculated with the total scores of the TULIA and the AST ratings performed at bedside and from video (rated by the third, independent rater, S.W.). With crosstabs, sensitivity, specificity, and negative (the probability that subjects with an AST test score of >9 truly lack gesture deficits in the full TULIA) and positive predictive values (the probability that subjects with an AST score of ≤9 truly have gesture deficits according to the full TULIA) were calculated for the AST cut-off score (see Tables 2, 3). We then applied Spearman's rho correlation on total scores as well as Cohen's kappa on a single-item level to test the association between the AST ratings from the videos (S.W.) and the same 12 items selected from the full TULIA (H.B.) to assess test-retest reliability. In order to exclude items with ceiling or floor effects, we checked the distribution of the ratings of TULIA items in patients. Displaying the frequencies on the single-item levels in histograms provided insight into the distribution of patient's gesture performance descriptively. This was expanded by calculating the items, for which more than 50% of the patients achieved values <5 and by scaling the variance from each patient to the mean. Two-tailed p values <0.05 were considered significant. All statistical tests were performed with "Statistical Package for Social Sciences" (SPSS), version 28 for windows. A detailed overview of the procedure is shown in Figure 1.

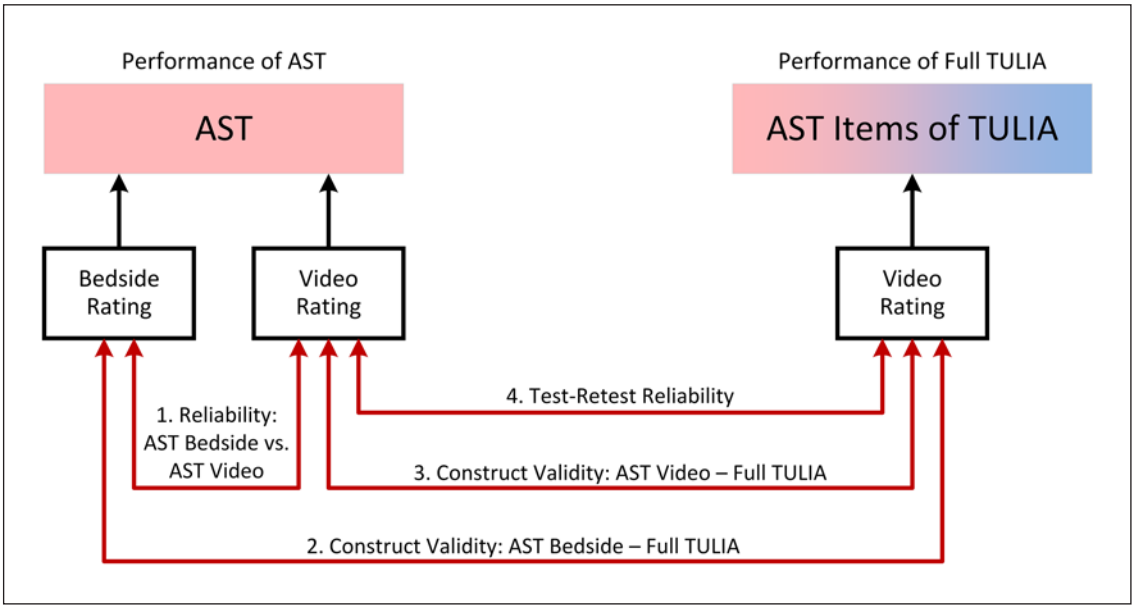


Fig. 1. Procedure of statistical tests.

Results

Clinical and demographic information is given in Table 1.

Reliability of AST Bedside versus AST Video

We found good parallel reliability between the AST ratings performed at bedside and from video (dichotomized) ($\rho = 0.869$; $p < 0.001$; $n = 27$). Inter-rater reliability at the individual item level reached a Cohen's kappa of 0.483 ($p < 0.001$; $n = 324$) between bedside and video AST ratings. Sensitivity was at 76.6%, while specificity reached 80.0%.

Construct Validity of AST Bedside

The total score of the AST rated at bedside (dichotomous rating) showed a significant correlation with the full TULIA on video ($\rho = 0.671$; $p < 0.001$; $n = 27$, see Fig. 2). Scatter plot in Figure 2 not only shows the correlation between the full TULIA and the AST, but also the cause for the moderate specificity of the AST. The performance of a few patients is just within the norm. Thus, according to the AST, they show gesture deficits.

The cut-off score for gesture deficits in patients with schizophrenia was set at 9 points (see Tables 2, 3 as well as Vanbellingen et al. [28], 2011). The diagnostic accuracy of the total score of the AST bedside compared to the full TULIA rated from video recordings showed good sensitivity (76.9%) and moderate specificity (50.0%) as previously seen in stroke patients. With a cut-off of 7 points, sensitivity declined to 61.5%, while specificity improved to 85.7%. With cut-offs above 9, the specificity decreased sharply.

Construct Validity of AST Video

The correlation of the AST rated from the videos (not dichotomized) with the full TULIA was strong ($\rho = 0.764$; $p < 0.001$; $n = 26$, see Fig. 3). Similar to the sensitivity and specificity, the positive and negative predictive value of the AST bedside improved when the AST rating was performed from videos (sensitivity of 83.3%, specificity of 64.3%, positive predictive value of 66.7%, negative predictive value of 81.8%). With a cut-off of 10, specificity decreased to 42.9% and sensitivity increased to 100%.

According to the AST bedside total score, 17 patients (63.0%), and, according to the video-rated AST total score, 15 patients (55.5%) had gesture deficits. In contrast, according to the total TULIA scores, 13 patients (48.1%) showed such deficits. There is no significant difference between the AST bedside mean score and the AST video mean score ($t(25) = -0.66$, p 0.511).

Fig. 2. Scatter plot showing the correlation between the full TULIA and the AST.

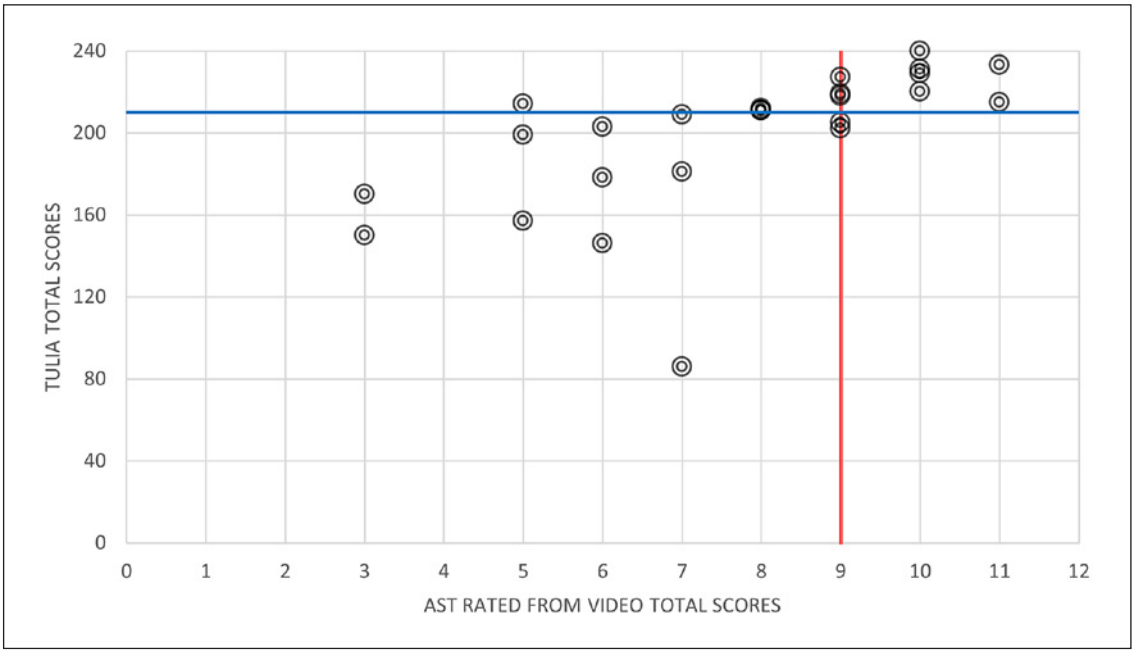
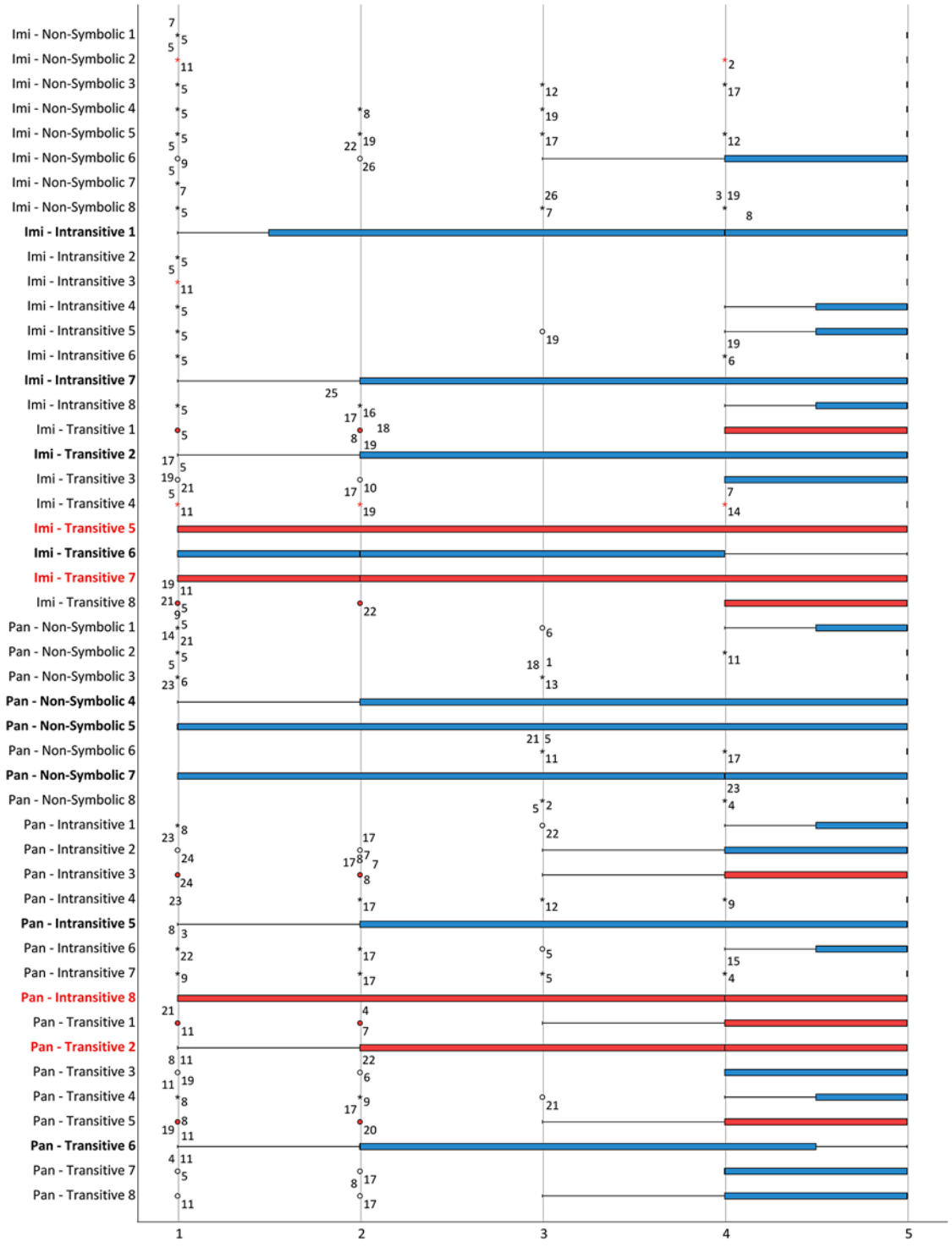


Fig. 3. Scatter plot showing the construct validity.

Fig. 4. Boxplot with the scores (0–5) for the 48 TULIA items of all 27 patients. Imi, Imitation, Pan, Pantomime. The red bars represent the 12 AST items, starting with no. 1 at the bottom of the boxplot. Note that AST item nos. 1, 2, and 4 form a ceiling effect, which is why only the numbers are marked with a red star and no bars are visible. The 13 TULIA items with sufficient variance are listed in bold black font, while red indicates the 4 items, which are included in the current AST. *(For figure see next page.)*



Imi = Imitation
Pan = Pantomime
The red bars represent the 12 AST items. Note that AST item nos. 1, 2 and 4 form a ceiling effect, which is why only the numbers are marked with a red star and no bars are visible.
The 13 TULIA items with sufficient variance are listed in bold black font, while red indicates the 4 items, which are included in the current AST

4

Reliability AST Video versus TULIA

As the AST rated from the videos strongly correlated with the full TULIA ($\rho = 0.764$; $p < 0.001$; $n = 26$), we explored test-retest reliability for specific items. The 12 AST items rated from videos (S.W.) and the same 12 items from the full TULIA (H.B.) strongly correlated regarding the total scores ($\rho = 0.772$; $p < 0.001$; $n = 26$), while the inter-rater agreement at the single-item level was fair (Cohens $\kappa = 0.312$; $p < 0.001$; $n = 312$). When the individual item values were dichotomized to pass or fail, the correlation between the 12 items of the AST and the TULIA, all rated from videos, was not substantially higher (Cohens $\kappa = 0.333$; $p < 0.001$; $n = 312$).

Explorative Analyses of TULIA Item Variance

For 79% of the TULIA items, more than half of the patients achieved the maximum score of 5, thus revealing little variance. Only in 13 of the 48 TULIA items, patients demonstrated a large variance, both when dichotomized (pass or fail) as well as when rated 0–5 (see Fig. 4, items in bold black font). The first six of these 13 items are from the imitation domain, and the latter seven are pantomime items. Critically, only four of the original 12 AST items are among those TULIA items with sufficient variation in schizophrenia (see Fig. 4, items in bold red font).

The internal consistency of the full TULIA shows a Cronbach α of 0.945 ($p < 0.001$; $n = 48$), leaving 5.5% of random variance. Similarly, the internal consistency of the AST items rated from video is 0.865 ($p < 0.001$; $n = 12$).

Discussion/Conclusion

Nonverbal communication skills are frequently compromised in schizophrenia [22], including the correct use of hand gestures. Identifying schizophrenia patients with gesture deficits is critical for outcome prediction [26]. The current TULIA procedure takes 45 min, including the assessment and rating from video recordings, while the full procedure of the AST at bedside requires, on average, 3 min. The extraordinary time efficiency for professionals and patients as well as the immediate availability of results at bedside are the strengths of the AST. Here, we tested the reliability of the AST in schizophrenia spectrum disorders, comparing it to the blinded ratings of the full TULIA.

The AST proves to be a reliable tool to assess gesture impairments in people with schizophrenia, as total AST scores from bedside and video ratings were strongly associated. This is in line with Vanbellingen et al. [28], where the AST and the TULIA score correlation was highly significant (Pearson's correlation coefficient $r = 0.96$, $p < 0.001$). Since in the current study only one examiner rated the patients from bedside, parallel reliability was calculated between the bedside and the video ratings to mimic inter-rater reliability. However, inter-rater reliability was only moderate [33] between online and video ratings of the AST at the single-item level. This is an unexpected contrast to the findings of Vanbellingen et al. [27] concerning the TULIA: Calculating inter-rater reliability with intra-class correlation coefficients and test-retest reliability with Spearman's rank correlation led to a good up to an excellent inter-rater reliability at an item level (range of weighted kappa's 0.65–0.99 and weighted percentage agreement of 80%) for the majority of items [28]. They excluded four items with lower inter-rater reliability (weighted kappa <0.60) from the 48 TULIA items for the AST [28]. Unfortunately, Vanbellingen et al. [28] did not survey inter-rater reliability of the AST on item basis. It remains unclear whether the AST for poststroke patients would also only allow moderate inter-rater reliability at a single-item level.

As Walther et al. [19, 20, 22, 24] (2013) suggested, with a higher cut-off score for schizophrenia patients in the TULIA, the threshold might also be higher in the AST than for poststroke patients (5 for severe, 9 for mild apraxia). Regarding the current sample, higher cut-off scores led to reduced specificity. Further studies are necessary to clarify this. In the current study, inter-rater reliability of the items could be compromised if the AST rating is performed by less experienced raters, in a heterogeneous sample of schizophrenia patients, or because of the different rating-forms (online vs. video, the latter providing more time to do the rating). Despite this moderate reliability on a single-item level, the total scores were highly consistent irrespective of the rating mode or examiners.

We detected good construct validity of the AST, as the full TULIA correlated significantly with the AST rated at bedside, and even stronger with the AST rated from video. However, the diagnostic accuracy of the AST bedside ratings was only moderate, with a better outcome regarding sensitivity, specificity, and positive and negative predictive values when applying AST video ratings. The current cut-off of the AST rated at bedside seems to be too sensitive for this population, resulting in many false positives. As a screening instrument, sensitivity and specificity should be above 80%. Lowering the cut-off to 7, the AST rated at bedside reached a more balanced specificity

of 86% and a sensitivity of 62%. Ideally, the AST had >80% sensitivity and 80–100% specificity. This would allow testing only those in whom a gesture deficit is suspected. However, gesture deficits are frequent in schizophrenia, and with increased sensitivity, a large number of subjects would have to be tested subsequently with the full TULIA (increasing workload). Therefore, we would rather aim for having close to 0 false negatives, i.e., increased specificity. This would allow running the full TULIA only in those identified as positive with the AST.

In comparison to the excellent sensitivity and specificity in stroke patients [28], the above-illustrated results in schizophrenia were clearly below our expectations. The moderate positive predictive value of 58.8% (number of true positives/sum of all positively tested) of the current AST rated at bedside indicates that positive results need retesting with the TULIA for further verification in order to confirm gesture deficits in clinical practice. This might also be important, as some features of gesture performance seem to be assessed more precisely when rated from video in contrast to bedside ratings. One explanation, therefore, might be that bedside rating needs more training, especially taking into account the reduced time available compared to the video ratings as in the full TULIA and the AST rated from video. Finally, in our sample of schizophrenia patients, the proportion of patients with gesture deficits varied substantially: the AST bedside test identified gesture deficits in 63.0% of the patients, while the video-rated AST (H.B.) identified gesture deficits in 37.0% and the full TULIA 48% of the patients. Collectively, these numbers are slightly lower than those published in previous studies [19, 22].

Besides the heterogeneous sample of schizophrenia patients, the item selection might provide explanations for the reduced predictive value of the AST. The item selection for the AST originally considered those suitable to detect apraxia in stroke patients [28] but might not divide well enough between schizophrenic subjects with or without a gesture deficit. In fact, gesture deficits are severe enough to be considered genuine apraxia in about 25% of the subjects suffering from schizophrenia [24].

Test-retest reliability between the AST rated from video by an independent rater and the full TULIA proved to be excellent on the level of the total scores. This is in line with the result achieved in stroke patients (AST ratings at bedside used) [28]. The correlation between the video-rated AST and the same TULIA items lost precision when analyzed on a single-item level. These correlations slightly improved when the scores were dichotomized as within the AST rated at bedside. Still, the inter-rater reliability of items that are shared by the AST and the TULIA remained fair. Similar to the above-mentioned lower inter-rater reliability at a single-item level between the AST at bedside and online, the rater's experience or heterogeneity of the sample could have influenced the AST's item performance. We suspect that the AST total score captures general gesture performance in schizophrenia fairly well, which is the main goal of its screening purpose.

Nearly 80% of the items from the full TULIA show limited variance in patients with schizophrenia. A significant proportion of them shows ceiling effects, i.e., more than 50% of the patients achieve the maximum score in these items. Only 13 TULIA items demonstrated sufficient variance among schizophrenia patients (see Fig. 4, items in bold black font). Of these 13 items, four are embedded in the current 12-item AST (see Fig. 4, items in bold red font). This might contribute to the reduced predictive value of the current AST in schizophrenia. When the AST was designed to detect apraxia in patients with brain lesions, one critical step of the item selection was the exclusion of TULIA items with low inter-rater reliability and internal consistency [28]. However, our findings suggest that this selection of items for the AST needs to be optimized for detecting gestural deficits in patients with schizophrenia. Consideration should be given to whether the other nine TULIA items with the highest variance (bold black font) should be included instead of the eight AST items with low variation. The current cut-off of 9 does not seem to be too liberal as cut-offs above 9 resulted in a decrease in specificity and only a slight increase in sensitivity. In the light of the small sample size and the variance, patients demonstrate between the AST items, it is challenging to derive meaningful novel cut-offs for schizophrenia. A revised AST for schizophrenia will need to test the above issues in a larger cohort. Importantly, while the AST is less effective in schizophrenia, our findings do not challenge the use of the AST in other types of apraxia.

This is the first study with the aim to validate the AST in schizophrenia. Clearly, larger patient samples are required in order to substantiate conclusions. However, an advantage of our patient sample is that it was nearly equally balanced regarding gender and included out- and inpatients. Furthermore, our sample consisted mainly of patients with long illness duration and moderate symptom severity. Most patients were administered antipsychotic medication, which could hamper gesture performance. Antipsychotic medication may differentially alter motor abnormalities, which is in turn strongly related to gesture performance in schizophrenia [20, 22, 24, 34].

However, previous studies failed to detect a medication effect on gesture performance [22]. At the time of our study, only 2 patients were off antipsychotic medication. No additional analyses were conducted by excluding them, as this subsample would also have been too small for any conclusions about medication effects. Finally, most gestures are culturally defined, allowing valid outcomes only for subjects who were raised within the corresponding cultural background. Our inclusion criteria met these concerns.

Subjects with schizophrenia often present impaired nonverbal skills, including poor gesture performance [17, 20, 22]. Critically, gesture deficits contribute to poor social function and quality of life [10, 12, 22, 26]. The gesture deficit in schizophrenia shares both the clinical and neurophysiological features of true apraxia, as in other neuropsychiatric disorders with impaired higher order motor control, such as Parkinson's disease [24]. The field is currently aiming at improving gesture skills in schizophrenia, applying multiple methods, including noninvasive brain stimulation and psychotherapy [35, 36]. However, these efforts would benefit from precise and fast identification of gesture deficits in clinical practice. Thus, the good construct validity found in our study between the full TULIA and the AST call for refining the selection of the TULIA items for a bedside psychosis-AST test. The predictive values could improve through this new item selection. On these grounds, it seems crucial to find an optimized item selection for a short test of TULIA. Rapid detection is required in order to allocate interventions to those with gesture deficits, as we may hope that interventions will not only improve gesture skills but will also exert downstream effects on social and occupational functioning [26].

In summary, our findings support the use of the AST to detect gesture deficits in subjects with schizophrenia. Furthermore, results call for refining the selection of the TULIA items for an improved bedside psychosis-AST test or to determine whether more experience would improve AST bedside ratings. The aim would be to increase specificity at reduced efforts. A psychosis-AST test may improve screening for gesture deficits in the clinics.

Acknowledgments

The authors would like to thank Thomas Hurni and Hallie Bachelet for their valuable inputs.

Statement of Ethics

All patients provided written informed consent. The study protocol adhered to the Declaration of Helsinki, 1975, had been approved by the cantonal Ethics Committee of Bern, KEK-BE (code: 025/2013). This investigation was part of a larger study that has been approved.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contributions

Hanta Bachofner performed measurements, i.e., rated the ASTs and the TULIAs from video, performed the analysis, interpreted the findings, and wrote the first draft of the manuscript. Konstantin A. Scherer recruited patients, conducted data acquisition, i.e., the examinations of the AST at bedside as well as the AST and the TULIA from video. He contributed to the interpretation of the findings. Sebastian Walther designed the study, supervised data acquisition, data analysis, and scientific writing. He performed parts of the measurements, contributed to data interpretation, and critically revised the manuscript. Tim Vanbellingen, Stephan Bohlhalter, and Katharina Stegmayer contributed to the interpretation of findings and critically reviewed the manuscript. All authors contributed to the final manuscript.

Data Availability Statement

The data are not publicly available due to legal restrictions.

References

1 Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1,160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the global burden of disease study 2010. Lancet. 2012 Dec 15;380(9859):2163–96.

2 Green MF, Horan WP, Lee J. Social cognition in schizophrenia. Nat Rev Neurosci. 2015; 16(10):620–31.

3 Ventura J, Hellemann GS, Thames AD, Koellner V, Nuechterlein KH. Symptoms as mediators of the relationship between neurocognition and functional outcome in schizophrenia: a meta-analysis. Schizophr Res. 2009 Sep; 113(2):189–99.

4 Lang FU, Kösters M, Lang S, Becker T, Jäger M. Psychopathological long-term outcome of schizophrenia – a review. Acta Psychiatr Scand. 2013;127(3):173–82.
5 Bellack AS, Green MF, Cook JA, Fenton W, Harvey PD, Heaton RK, et al. Assessment of community functioning in people with schizophrenia and other severe mental illnesses: a white paper based on an NIMH-sponsored workshop. Schizophr Bull. 2007; 33(3):805–22.
6 Velthorst E, Fett AJ, Reichenberg A, Perlman G, van Os J, Bromet EJ, et al. The 20-year longitudinal trajectories of social functioning in individuals with psychotic disorders. Am J Psychiatry. 2017 Nov 1;174(11):1075–85.
7 Burns T, Patrick D. Social functioning as an outcome measure in schizophrenia studies. Acta Psychiatr Scand. 2007;116(6):403–18.
8 Toomey R, Schuldberg D, Corrigan P, Green MF. Nonverbal social perception and symptomatology in schizophrenia. Schizophr Res. 2002 Jan;53(1):83–91.
9 Bucci S, Startup M, Wynn P, Baker A, Lewin TJ. Referential delusions of communication and interpretations of gestures. Psychiatry Res. 2008 Feb 28;158(1):27–34.
10 White TP, Borgan F, Ralley O, Shergill SS. You looking at me?: interpreting social cues in schizophrenia. Psychol Med. 2016;46(1): 149–60.
11 Park S, Matthews N, Gibson C. Imitation, simulation, and schizophrenia. Schizophr Bull. 2008;34(4):698–707.
12 Walther S, Mittal VA. Why we should take a closer look at gestures. Schizophr Bull. 2016; 42(2):259–61.
13 Obermeier C, Dolk T, Gunter TC. The benefit of gestures during communication: Evidence from hearing and hearing-impaired individuals. Cortex. 2012 Jul;48(7):857–70.
14 Kircher T, Straube B, Leube D, Weis S, Sachs O, Willmes K, et al. Neural interaction of speech and gesture: differential activations of metaphoric co-verbal gestures. Neuropsychologia. 2009;47(1):169–79.
15 Weiss PH, Ubben SD, Kaesberg S, Kalbe E, Kessler J, Liebig T, et al. Where language meets meaningful action: a combined behavior and lesion analysis of aphasia and apraxia. Brain Struct Funct. 2016;221(1):563–76.
16 Troisi A, Spalletta G, Pasini A. Non-verbal behaviour deficits in schizophrenia: an ethological study of drug-free patients. Acta Psychiatr Scand. 1998;97(2):109–15.
17 Mittal VA, Tessner KD, McMillan AL, Delawalla Z, Trotman HD, Walker EF. Gesture behavior in unmedicated schizotypal adolescents. J Abnorm Psychol. 2006;115(2):351.
18 Lavelle M, Healey PG, McCabe R. Is nonverbal communication disrupted in interactions involving patients with schizophrenia? Schizophr Bull. 2013;39(5):1150–8.
19 Walther S, Vanbellingen T, Müri R, Strik W, Bohlhalter S. Impaired gesture performance in schizophrenia: particular vulnerability of meaningless pantomimes. Neuropsychologia. 2013;51:2674–8.
20 Walther S, Vanbellingen T, Müri R, Strik W, Bohlhalter S. Impaired pantomime in schizophrenia: association with frontal lobe function. Cortex. 2013;49(2):520–7.
21 Millman ZB, Goss J, Schiffman J, Mejias J, Gupta T, Mittal VA. Mismatch and lexical retrieval gestures are associated with visual information processing, verbal production, and symptomatology in youth at high risk for psychosis. Schizophr Res. 2014 Sep;158(1–3):64–8.
22 Walther S, Stegmayer K, Sulzbacher J, Vanbellingen T, Müri R, Strik W, et al. Nonverbal social communication and gesture control in schizophrenia. Schizophr Bull. 2015 Mar; 41(2):338–45.
23 Strik W, Wopfner A, Horn H, Koschorke P, Razavi N, Walther S, et al. The Bern psychopathology scale for the assessment of system-specific psychotic symptoms. Neuropsychobiology. 2010;61(4):197–209.
24 Walther S, Mittal VA, Stegmayer K, Bohlhalter S. Gesture deficits and apraxia in schizophrenia. Cortex. 2020;133:65–75.
25 Matthews N, Gold BJ, Sekuler R, Park S. Gesture imitation in schizophrenia. Schizophr Bull. 2013 Jan;39(1):94–101.
26 Walther S, Eisenhardt S, Bohlhalter S, Vanbellingen T, Müri R, Strik W, et al. Gesture performance in schizophrenia predicts functional outcome after 6 months. Schizophr Bull. 2016;42(6):1326–33.
27 Vanbellingen T, Kersten B, Van Hemelrijk B, Van de Winckel A, Bertschi M, Müri R, et al. Comprehensive assessment of gesture production: a new test of upper limb apraxia (TULIA). Eur J Neurol. 2010 Jan;17(1):59–66.
28 Vanbellingen T, Kersten B, Van de Winckel A, Bellion M, Baronti F, Müri R, et al. A new bedside test of gestures in stroke: the apraxia screen of TULIA (AST). J Neurol Neurosurg Psychiatry. 2011 Apr;82(4):389–92.
29 Stegmayer K, Moor J, Vanbellingen T, Bohlhalter S, Müri RM, Strik W, et al. Gesture performance in first-and multiple-episode patients with schizophrenia spectrum disorders. Neuropsychobiology. 2016;73(4):201–8.
30 Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia. 1971;9(1):97–113.
31 Woods SW. Chlorpromazine equivalent doses for the newer atypical antipsychotics. Journal of Clinical Psychiatry. 2003;64(6):663–667.
32 De Renzi E, Motti F, Nichelli P. Imitating gestures. A quantitative approach to ideomotor apraxia. Arch Neurol. 1980 Jan;37(1):6–10.
33 Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159–74.
34 Peralta V, Cuesta MJ. The effect of antipsychotic medication on neuromotor abnormalities in neuroleptic-naive nonaffective psychotic patients: a naturalistic study with haloperidol, risperidone, or olanzapine. Prim Care Companion J Clin Psychiatry. 2010; 12(2):e1–11.
35 Lefebvre S, Pavlidou A, Walther S. What is the potential of neurostimulation in the treatment of motor symptoms in schizophrenia? Expert Rev Neurother. 2020 Jul 2;20(7):697–706.
36 Walther S, Alexaki D, Schoretsanitis G, Weiss F, Vladimirova I, Stegmayer K, et al. Inhibitory repetitive transcranial magnetic stimulation to treat psychomotor slowing: a transdiagnostic, mechanism-based randomized double-blind controlled trial. Schizophr Bull Open. 2020;1(1).