Comprehensive assessment of gesture production: A new test of upper limb apraxia (TULIA)

Article *in* European Journal of Neurology · August 2009

DOI: 10.1111/j.1468-1331.2009.02741.x · Source: PubMed

Tim Vanbellingen
Luzerner Kantonsspital

Bernd Kersten
Universität Bern

Ann Van de Winckel
University of Minnesota Twin Cities

Manuel Bertschi
neuroaarau AG

European Journal of Neurology 2010, **17:** 59–66 doi:10.1111/j.1468-1331.2009.02741.x

Comprehensive assessment of gesture production: a new test of upper limb apraxia (TULIA)

T. Vanbellingen[a], B. Kersten[b], B. Van Hemelrijk[a], A. Van de Winckel[c], M. Bertschi[d], R. Müri[d], W. De Weerdt[c] and S. Bohlhalter[d]

[a]*Klinik Bethesda, Neurorehabilitation Centre, Tschugg, Switzerland;* [b]*Department of Psychology, University of Bern, Bern, Switzerland;* [c]*Department of Rehabilitation Sciences, K.U. Leuven, Belgium; and* [d]*Division of Cognitive and Restorative Neurology, Department of Neurology, Inselspital, University Hospital Bern, Bern, Switzerland*

Keywords:
apraxia, imitation, pantomime, reliability, scale, validity

Received 2 March 2009
Accepted 5 June 2009

Background: Only few standardized apraxia scales are available and they do not cover all domains and semantic features of gesture production. Therefore, the objective of the present study was to evaluate the reliability and validity of a newly developed test of upper limb apraxia (TULIA), which is comprehensive and still short to administer.

Methods: The TULIA consists of 48 items including imitation and pantomime domain of non-symbolic (meaningless), intransitive (communicative) and transitive (tool related) gestures corresponding to 6 subtests. A 6-point scoring method (0–5) was used (score range 0–240). Performance was assessed by blinded raters based on videos in 133 stroke patients, 84 with left hemisphere damage (LHD) and 49 with right hemisphere damage (RHD), as well as 50 healthy subjects (HS).

Results: The clinimetric findings demonstrated mostly good to excellent internal consistency, inter- and intra-rater (test–retest) reliability, both at the level of the six subtests and at individual item level. Criterion validity was evaluated by confirming hypotheses based on the literature. Construct validity was demonstrated by a high correlation ($r = 0.82$) with the De Renzi-test.

Conclusion: These results show that the TULIA is both a reliable and valid test to systematically assess gesture production. The test can be easily applied and is therefore useful for both research purposes and clinical practice.

Introduction

Limb apraxia is defined by the inability to correctly perform skilled and/or learned limb movements, which cannot be explained by elementary motor and sensory deficits, or cognitive problems [1]. Praxis movements can be classified in two main categories: gestures and real tool/object use. Most clinical tests on apraxia focus on the examination of gestures, whose impairment has been demonstrated to be predictive for difficulties in actual tool use [2,3]. Gestures can be tested in two principal domains, *imitation* of gestures demonstrated by the examiner and *pantomime* on command (elicited verbally), as well as for three types of their semantic content. Accordingly, gestures are *non-symbolic* if they are meaningless for the examinee (e.g. index finger on top of nose) and symbolic if they are meaningful. Symbolic gestures are further classified as being *intransitive*, communicative in nature (e.g. military salute) or being *transitive*, related to specific object/tool use (e.g. hammer use).

There is clinical evidence that apraxic patients may be differentially affected for these domains and semantic aspects. Patients with stroke, particularly if left hemispheric, show typically more difficulties pantomiming than imitating gestures and performing transitive than intransitive or meaningless movements [4]. Furthermore, recent brain mapping studies suggest that their neural representation may be distinct [5–8].

Apraxia is increasingly recognized as a cognitive-motor disorder of stroke with considerable impact on everyday life [9–11]. It is associated with poorer outcome for independent living [12,13] or return to work [14]. Therefore, a reliable and valid assessment of apraxia, which is not too time-consuming, is important for both clinical evaluation and research protocols [15]. Currently, the available scales [16–27] either do not

Correspondence: Stephan Bohlhalter, Division of Cognitive and Restorative Neurology, Department of Neurology, Inselspital, Bern University Hospital, Switzerland (tel.: + 41 31 632 90 97; fax: + 41 31 632 97 70; e-mail: stephan.bohlhalter@insel.ch).

meet standard clinimetric criteria or do not cover all relevant behavioral and motor aspects of gesture production. Accordingly, reliability was assessed only in four studies [19,25–27] and validity in none of the reports. Furthermore, only two studies [17,24] considered all domains and semantic categories, but provided no clinimetric assessment. Finally, many testing instruments lack sensitivity in the scoring for clinical changes and/or disregard qualitative information on how movements are disturbed. Therefore, we aimed to develop a novel test for upper limb apraxia (TULIA), fulfilling all the aforementioned requirements.

Methods

Sample

A cohort of 133 patients with a single unilateral stroke (ischemic or hemorrhagic), 84 with left hemispheric damage (LHD), 49 with right hemispheric damage (RHD), and 50 healthy subjects (HS) participated in this study. If categorized according to the localization within hemispheres, lesions involved cortical damage (CD) in 96 patients, or were restricted to subcortical damage (SCD) in 37 patients. SCD patients had small strokes, mainly in striatocapsular localization. The patients were admitted to five different neurological rehabilitation centres, three in Switzerland (Klinik Bethesda, Tschugg, Inselspital Bern, and Klinik Valens) and two in Germany (Kiliani-Klinik, Bad Windsheim and Fachklinik Herzogenaurach). Patients were included if the stroke was radiologically confirmed by CT or MRI. Exclusion criteria were other brain disorders, particularly dementia and Parkinson's disease. Furthermore, patients with orthopaedic disorders of the upper non-hemiplegic limb that could interfere with the TULIA task were excluded. Healthy subjects were included if free of any history of neurological disease. All subjects had normal corrected sight and hearing, and spoke German as native language. Written informed consent was obtained in all subjects according to the Declaration of Helsinki, 1975. No significant differences were found for demographic and clinical data between subjects as listed in Table 1.

Table 1 Characteristics of patients and healthy subjects [mean (SD)][a]

	Hemisphere damage		
Characteristic	Left (LHD) *n* (%)	Right (RHD) *n* (%)	HS *n* (%)
Total	84	49	50
Age	63.51 (12.49)	65.63 (10.98)	61.46 (10.99)
Range (years)	30–84	33–85	43–93
Gender			
Women	30 (35.7)	16 (32.7)	33 (66)
Men	54 (64.3)	33 (67.3)	17 (34)
Type of stroke			
Haemorrhage	13 (15.5)	7 (14.3)	
Infarction	71 (84.5)	42 (85.7)	
Lesion Location			
Cortical (CD)	63 (75)	33 (67.3)	
Subcortical (SCD)	21 (25)	16 (32.7)	
Time-interval after stroke			
Subacute (> 3 days)	65 (77.4)	40 (81.6)	
Chronic (> 3 months)	19 (22.6)	9 (18.4)	
Time post-onset to initial testing			
Period of days	80.77 (112.77)	110.96 (296.36)	
Aphasia	59 (70.2)	0	

LHD, left hemisphere damage; RHD, right hemisphere damage; CD, cortical damage; SCD, subcortical damage; HS, healthy subjects.
[a]No significant difference was found between subjects with respect to age ($F = 1.57$, $P > 0.05$), sex ($\chi^2 = 0.13$, $P = 0.72$), type of stroke ($\chi^2 = 0.03$, $P = 0.85$), lesion localization ($\chi^2 = 1.43$, $P = 0.49$), time-interval after stroke ($\chi^2 = 0.34$, $P = 0.56$), and time (days) post-onset to initial test of upper limb apraxia (TULIA) ($F = 3.03$, $P > 0.05$).

Scale development: item selection and scoring method

The concept of the TULIA test was developed along the principal domains and semantic aspects known for gesture production. Hence, the scale consisted of 48 items incorporating six subtests for both imitation and pantomime of non-symbolic, intransitive and transitive gestures (see appendix 1). The items were compiled from various sources [18,20,21,24,26,28,29]. They were selected based on their kinematic features such as proximal and distal as well as simple and repetitive movements. Items were also selected based on clinical experience, that is, on gestural difficulties observed in many apraxic patients.

A 6-point scoring method was developed (see Appendix 2). For each item the minimum score was zero and the maximum score was five; therefore, the potential score range was 0–240. The scoring system followed a stepwise judgement. Firstly, achievement of overall movement goal was evaluated, narrowing the range of scores to either 0 and 1, or 2–5. The goal was considered not achieved if trajectories of gestures were grossly disturbed or if, in the case of symbolic gestures, their semantic content was incorrect. Trajectories reflect the spatial orientation of movements including movement plane relative to goal object (tool or own body), joint coordination and movement shape (e.g. chopping or sliding). Secondly, detailed error analysis (e.g. body part as object errors) according to widely used classifications [20,24] allowed final scoring, particularly in the upper range (2–5).

Procedures

The TULIA test required about 20 min to administer, including preparing the subject's position and giving the necessary instructions. Subjects were seated in front of the examiner; both with the forearms placed on the table. Patients executed the movements with their non-paretic upper limb. Healthy subjects used their left hand only since previous studies showed no difference between right or left hand performance [18,30]. The examiner demonstrated the movements in a mirrored fashion. Subjects were instructed to reproduce the movement after the examiners completed their demonstration. On pantomime, subjects were asked to carefully listen to the task and to execute it as precisely as possible. Examiners were using the formulae in appendix 1. To maintain attention, subjects were informed whether the movements had a meaning or not before each subtest. For transitive gestures subjects were explicitly instructed to imagine holding a tool/object in their hand. The sequence of subtests was fixed (according to appendix 1). All subjects' performances were videotaped for later scoring. Three experienced examiners scored the videos independently: T.V. and B.H. for inter-rater reliability, T.V. and F.A. (see acknowledgements) for test-retest reliability and construct validity (TULIA and De Renzi test [18]).

Statistical analysis

Statistical analyses were performed using SPSS for Windows (Version 15.0.0; SPSS, Inc. Chicago, IL, USA) and SAS (Version 8.2, SAS, Institute Inc., Cary, NC, USA). For the summed scores of subtests internal consistency was assessed by Cronbach's α, inter-rater reliability using intra-class correlation coefficients and test–retest reliability with Spearman's rank correlation analysis [31]. In addition, at individual item level, inter-rater and test–retest-reliability was calculated using weighted kappa statistics and weighted percentage agreement [32].

One-way ANOVA was used for the analysis of criterion validity between subject's differences of total TULIA scores. Furthermore, we performed a multivariate ANOVA (MANOVA) with the summed scores of the subtests (domains and semantic features) as dependent variables and subject's groups (LHD vs. RHD vs. healthy subjects) as between-subject factor. Differences of interest were calculated by *post hoc* Tukey honestly significant difference (HSD) test. Within subject's analysis (both between and within domains) was examined using repeated measures ANOVA. For evaluation of the construct validity the total score of TULIA was compared with the total score of De Renzi test [18] by Spearman's rank correlation analysis.

Results

Frequency and severity of apraxia in the patient cohort

There was no ceiling or floor effect as the lowest and maximum scores were not reached by any subject. Based on a cut-off score of 194 (two SD below mean score of healthy subjects) within the LHD group 13% patients were categorized as severely apraxic (< 65, one-third of cut-off), 12% as moderately (< 130, two-third of cut-off) and 43% as mildly apraxic [16]. In the RHD group none of the patients was severely, only 2% were moderately and 37% mildly apraxic. Patients with cortical lesions were more frequently apraxic, particularly if severely (10.4% vs. 2.7%) or moderately affected (9.4% vs. 5.4%).

Internal consistency, inter- and intra-rater (test–retest) reliability

Internal consistency and inter-rater reliability was assessed in 52 patients and 12 healthy subjects, intra-rater reliability (test–retest) was assessed in 20 patients (examined three times within 24 h). The findings (Table 2) indicate high consistency and reproducibility for both domains and most semantic aspects, being partly lower only for imitation of non-symbolic gestures.

For inter-rater reliability at item level the majority of items had good ($n = 12$) to almost perfect ($n = 32$) degree of agreement (range of weighted kappa's 0.65–0.99). The remaining items ($n = 4$) showed fair (item 34 and 46) to moderate (item 3 and 48) agreement (range 0.35–0.50). For test–retest reliability at item level most items showed good ($n = 10$) to almost perfect ($n = 29$) degree of agreement (range 0.66–1.0), being fair ($n = 5$) to moderate ($n = 3$) for the rest (0.30–0.60).

Table 2 Internal consistency, intraclass correlation and test-retest reliability of subtests

	Cronbach's α^a	ICC^b	Test–retest^c Tests (1–2)	Tests (1–3)	Tests (2–3)
Imitation total	0.93	0.96	0.85	0.86	0.93
Non-symbolic	0.67	0.90	0.66	0.46	0.78
Intransitive	0.83	0.91	0.74	0.79	0.89
Transitive	0.87	0.95	0.85	0.91	0.94
Pantomime total	0.96	0.99	0.94	0.95	0.98
Non-symbolic	0.87	0.98	0.88	0.92	0.93
Intransitive	0.90	0.89	0.93	0.95	0.93
Transitive	0.91	0.96	0.91	0.96	0.94

[a]Cronbach's alpha > 0.7 = satisfactory.
[b]Intraclass correlation coefficients > 0.75 = excellent.
[c]Spearman's rank correlation coefficients $\rho > 0.7$ = good.

Criterion validity

Criterion-related validity of the TULIA test was assessed by exploring its ability to confirm hypotheses based on the widely accepted clinical fact that impairment of gesture production represents primarily a cortical syndrome, which is more common and severe in left than right hemispheric stroke. Accordingly, amongst LDH, RHD and healthy subject groups significant differences were found ($F = 36.319$, $P < 0.001$, $\eta = 0.29$), with significantly lower total TULIA scores of LHD than RHD patients (154.9 ± 59.2 and 193.4 ± 23.8, $P < 0.001$, *post hoc* Tukey HSD) and less pronounced of RHD patients than healthy subjects (193.4 ± 23.8 and 217.5 ± 11.7, $P < 0.05$) (Fig. 1a. Similar effects were found between CD, SCD and healthy subjects ($F = 25.484$, $P < 0.001$, $\eta = 0.22$), total scores of CD patients being significantly lower than those of SCD patients (162.35 ± 55.5 and 186.49 ± 39.4, $P < 0.05$) and of SCD patients lower than of healthy subjects (186.49 ± 39.4 and 217.5 ± 11.7, $P < 0.05$) (Fig. 1b).

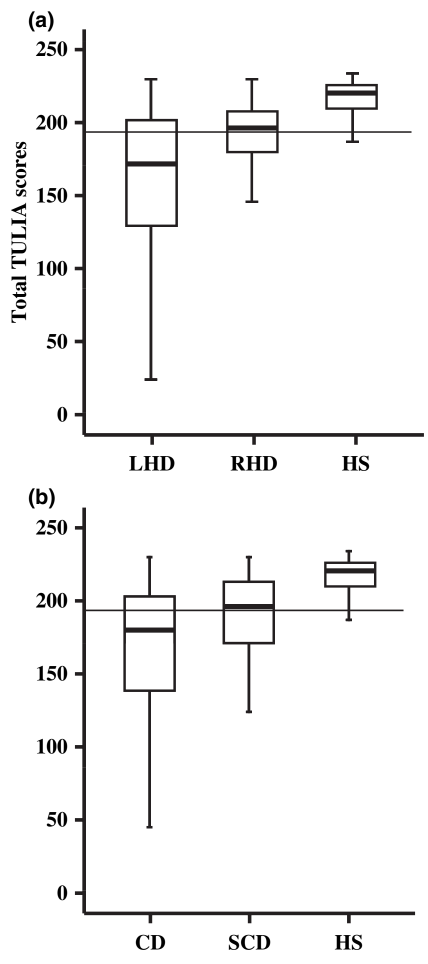

Figure 1 Boxplots (with lowest and highest scores as well as median separating lower and upper quartile) displaying differences of total test of upper limb apraxia (TULIA) scores between left hemisphere damage, right hemisphere damage and healthy subjects (a) and cortical damage, subcortical damage and healthy subjects (b). Solid lines indicate the cut-off score of 194 for apraxia (two SD below mean score of healthy subjects).

At the level of subtests (summarized in Table 3), scores were significantly lower in LHD than RHD patients, regardless of domain. By contrast, significant differences between RHD and healthy subjects were found predominantly in the imitation domain. Within patients with LHD overall performance in pantomime was significantly more difficult than imitation of gestures, which was not detected for the RHD patients and healthy subjects. Furthermore, between domains significant differences were found for non-symbolic gestures in all groups, which probably accounts for difference in overall pantomime measures of LHD. By contrast, pantomime performances of symbolic gestures were either superior, particularly for intransitive gestures in RHD or did not show any difference. Analysis within domains showed that performance of transitive gestures was generally more difficult than intransitive gestures for all groups. Similarly, transitive gestures were more demanding than non-symbolic gestures, however, only in the imitation domain.

Construct validity

Thirty-three patients (21 LHD and 12 RHD) were tested with both TULIA and De Renzi test [18]. Correlation between TULIA and De Renzi scores was strong ($r = 0.82$) indicating that the two scales measured related constructs. Furthermore, because the De Renzi test only includes gestures performed on imitation separate correlation analysis was also performed for the two subscales. A higher correlation was found for the imitation subscale and De Renzi test ($\rho = 0.80$) than for the pantomime subscale ($\rho = 0.65$).

Discussion

Test for upper limb apraxia is a test primarily of gesture production, a process involving the recruitment of time-space programs of skilled movements and their translation into proper limb trajectories. Impaired gesture production is the hallmark of so called ideomotor apraxia, the most frequent apraxia disorder in stroke [1]. The conceptual component of praxis processing is embedded in the scoring system as content errors are registered, particularly in lower score ranges. The TULIA test is comprehensive since all relevant behavioural (represented in the subtests) and kinematic aspects of gesture production are considered, which

Table 3 Mean (SD) TULIA scores of LHD, RHD and HS

	LHD	P^a	RHD	P^a	HS	Cut-off
Total score	154.9 (59.2)	< 0.05	193.4 (23.8)	< 0.05	217.5 (11.7)	194
Imitation	80.3 (25.0)[b]	< 0.05	95.4 (14.3)	< 0.05	108.1 (6.3)	95
Non-symbolic	30.9 (6.6)[b]	< 0.05	33.9 (4.3)[b]	> 0.05	36.7 (2.7)	31
Intransitive	26.4 (9.4)[b]	< 0.05	31.8 (5.5)[b]	< 0.05	37.0 (2.9)	31
Transitive	23.0 (10.9)	< 0.05	29.6 (6.9)[b]	> 0.05	34.3 (3.3)	28
Pantomime	74.2 (36.1)[b]	< 0.05	98.0 (13.8)	> 0.05	108.7 (8.2)	92
Non-symbolic	22.9 (11.8)[b]	< 0.05	29.7 (6.1)[b]	< 0.05	35.4 (3.9)	28
Intransitive	28.3 (13.4)[b]	< 0.05	36.2 (4.4)[b]	> 0.05	37.9 (2.8)	32
Transitive	23.0 (12.6)	< 0.05	32.2 (6.6)[b]	> 0.05	35.3 (4.7)	26

LHD, left hemisphere damage; RHD, right hemisphere damage, HS, healthy subjects. [a]Between subject's differences (MANOVA, Bonferroni corrected, *post hoc* tests).
[b]Within subject's differences, between domains, $P < 0.05$ (repeated measures ANOVA, Bonferroni corrected). Within domains all $P < 0.05$, except for non-symbolic vs. transitive within pantomime of LHD and HS $P > 0.05$.

may be differentially affected in patients with apraxia [1]. TULIA allows systematic investigation of these dissociations, which were assessed by non-validated or incomplete scales, so far (see Table 4).

Although frequency of mild apraxia was similar for both stroke groups, severer apraxia was, as expected, almost only found in the LHD group. Furthermore, apraxia was also more prevalent and severe in CD than SCD patients [33,34].

The present study demonstrated that the summed scores of the TULIA subscales were highly consistent and reproducible, except for the imitation of non-symbolic gestures, which was in part less reliable. Furthermore, good to excellent inter-rater agreement was also obtained at the item level as most items were above weighted kappa values of 0.60 and weighted percentage agreement of 80%. Only four items showed fair to moderate agreement. Similarly, test–retest reliability at item level demonstrated good to excellent reproducibility for 40 items, being lower only for eight items (slight to moderate agreement). High intra-rater reliability is particularly important for research purposes, making the TULIA test a suitable instrument for intervention studies. The high internal consistency points to some redundancy of the TULIA items. However, to maintain the theoretical structure of the test, we did not perform an item reduction, which probably would have been possible following strict clinimetric principles.

Validation of the TULIA test was performed by assessing its ability to reflect external criteria, that is, to confirm several straightforward hypotheses originating from clinical knowledge and current literature. Specifically, the main hypothesis was that in patients with LHD, in whom apraxia is typically more frequent and severe [4], as also shown herein, total scores of TULIA would be significantly lower than the RHD. It was also expected that patients with CD would score lower than SCD. The findings demonstrate that the TULIA test was able to fulfil these external criteria. In addition, since apraxic

Table 4 Comparison of existing apraxia tests for gesture production in stroke patients (since 1980)

	Reliability (inter-rater or test-retest)	Validity (construct or criterion)	Transitive gestures	Intransitive gestures	Meaningless gestures	Domain tested	Scoring	Cut-off scores
De Renzi *et al.* [18]	–	–	+	+	+	Imitation only	0–3	+
Kertesz *et al.* [16]	–	–	+	+	–	Pantomime and imitation	0–3	+
Poeck [24]	–	–	+	+	+	Pantomime and imitation	fail or pass	–
Alexander *et al.* [26]	+	–	+	+	–	Pantomime and imitation	0–5	+
Goldenberg [38]	–	–	–	–	+	Imitation only	1–2	–
Rothi *et al.* [20]	–	–	+	+	–	Pantomime only	fail or pass	+
Dobigny-Roman *et al.* [19]	+	–	–	+	+	Imitation only	0–3	–
Roy *et al.* [28]	–	–	+	–	–	Pantomime and imitation	0–3	+
van Heugten [39], Zwinkels *et al.* [23]	+	–	+	+	–	Pantomime and imitation	summed scores	+
Bartolo *et al.* [17]	–	–	+	+	+	Pantomime and imitation	fail or pass	+
TULIA	+	+	+	+	+	Pantomime and imitation	0–5	+

TULIA, test of upper limb apraxia.
Available (+) and not available (–).

errors are known to occur in RHD [35] and SCD [33] patients, the hypothesis that they score lower in the TULIA test than the healthy was also verified. We did not formally test specificity and sensitivity since a gold standard measure of apraxia or a reliable clinical assessment is not available. However, the significant differences between stroke patients groups and healthy subjects indicate that the TULIA test is able to distinguish between apraxic and non-apraxic subjects. We determined cut-off scores for all subtests, which allow assessing the differential patterns of apraxia.

Pantomime on verbal command offers few cues in contrast to imitation, in which the gestural input provides a model for the correct performance. Therefore, pantomime is usually more difficult for patients than imitation, particularly with LDH. Our findings confirmed this premise with pantomime scores being significantly lower than imitation scores in patients with LHD. However, this seems to be related mainly to the difference in performance of non-symbolic gestures that pose particular demands on gesture planning, since patients cannot rely on a gestural model provided by the examiner or on learned engrams supporting meaningful gestures. The reason for the lack of difference between imitation and pantomime (in LHD) or even better performance in pantomime (in RHD) of symbolic gestures remains unclear. It is conceivable that some practice effect may play a role, as imitation was always tested before pantomime. Transitive movements are cognitively more complex as their performance involves an imagined interaction with a tool. Our findings that transitive subscores were significantly lower than intransitive subscores, for patients as well as healthy subjects and irrespective of involved domains, were therefore anticipated [36].

A high correlation between the TULIA and an existing apraxia test was found (ρ = 0.82 with De Renzi test), indicating that related constructs of gesture production were measured. As expected, correlation was lower for the pantomime than the imitation subscale, as the de Renzi test assesses imitation only. Validity is further supported by the finding that few patients apraxic in the TULIA test were not apraxic in the De Renzi test and vice versa. We choose the De Renzi test since it is evaluated with cut-off scores determined in a large population [18], although in contrast to the De Renzi test, the items of TULIA were not mixed within subtests allowing better dissociation between subcategories [37].

Conclusion

The most important advantage of the TULIA test is that it allows a comprehensive evaluation of gesture production, which is yet short to administer. Furthermore, the TULIA test fulfils clinimetric standards such as high reliability and validity, which is indispensable, for instance, in apraxia treatment protocols [9]. In addition, the TULIA test provides a detailed scoring system, which is sensitive for clinical changes. The test can be scored instantaneously by experienced raters. However, as video monitoring is considered the gold standard even for clinical use [20], the reliability was assessed for video based rating in this study. The TULIA test has not been conceptualized for bedside evaluation. Therefore, based on the present work, a screening version of the TULIA will be developed.

Acknowledgements

The authors wish to thank Kiliani-Klinik (Germany), Fachklinik Herzogenaurach (Germany), and Inselspital Bern (Switzerland) for their support and accommodation. We also thank Flavia Antonietti for her crucial assistance in evaluating the videotapes. We are grateful to Dr. Brian Martin for skilful editing of the manuscript.

References

1. Ochipa C, Gonzalez Rothi LJ. Limb apraxia. *Semin Neurol* 2000; **20:** 471–478.
2. Goldenberg G, Hagmann S. Tool use and mechanical problem solving in apraxia. *Neuropsychologia* 1998; **36:** 581–589.
3. Poizner H, Clark MA, Merians AS, Macauley B, Gonzalez Rothi LJ, Heilman KM. Joint coordination deficits in limb apraxia. *Brain* 1995; **1:** 227–242.
4. Leiguarda RC, Marsden CD. Limb apraxias: higher-order disorders of sensorimotor integration. *Brain* 2000; **123:** 860–879.
5. Bohlhalter S, Hattori N, Wheaton L, *et al*. Gesture subtype-dependent left lateralization of praxis planning: an Event-Related fMRI Study. *Cereb Cortex* 2009; **19:** 1256–1262.
6. Buxbaum LJ, Kyle K, Grossman M, Coslett HB. Left inferior parietal representations for skilled hand-object interactions: evidence from stroke and corticobasal degeneration. *Cortex* 2007; **43:** 411–423.
7. Goldenberg G, Hermsdorfer J, Glindemann R, Rorden C, Karnath HO. Pantomime of tool use depends on integrity of left inferior frontal cortex. *Cereb Cortex* 2007; **17:** 2769–2776.
8. Goldenberg G, Karnath HO. The neural basis of imitation is body part specific. *J Neurosci* 2006; **26:** 6282–6287.
9. Buxbaum LJ, Haaland KY, Hallett M, *et al*. Treatment of limb apraxia: moving forward to improved action. *Am J Phys Med Rehabil* 2008; **87:** 149–161.
10. Hanna-Pladdy B, Heilman KM, Foundas AL. Ecological implications of ideomotor apraxia: evidence from physical activities of daily living. *Neurology* 2003; **60:** 487–490.
11. Sunderland A, Shinner C. Ideomotor apraxia and functional ability. *Cortex* 2007; **43:** 359–367.

12. Sundet K, Finset A, Reinvang I. Neuropsychological predictors in stroke rehabilitation. *J Clin Exp Neuropsychol* 1988; **10:** 363–379.
13. Wetter S, Poole JL, Haaland KY. Functional implications of ipsilesional motor deficits after unilateral stroke. *Arch Phys Med Rehabil* 2005; **86:** 776–781.
14. Saeki S, Hachisuka K. The association between stroke location and return to work after first stroke. *J Stroke Cerebrovasc Dis* 2004; **13:** 160–163.
15. Wheaton LA, Hallett M. Ideomotor apraxia: a review. *J Neurol Sci* 2007; **260:** 1–10.
16. Kertesz A, Ferro JM. Lesion size and location in ideomotor apraxia. *Brain* 1984; **3:** 921–933.
17. Bartolo A, Cubelli R, Sala SD. Cognitive approach to the assessment of limb apraxia. *Clin Neuropsychol* 2008; **22:** 27–45.
18. De Renzi E, Motti F, Nichelli P. Imitating gestures. A quantitative approach to ideomotor apraxia. *Arch Neurol* 1980; **37:** 6–10.
19. Dobigny-Roman N, Dieudonne-Moinet B, Tortrat D, Verny M, Forette B. Ideomotor apraxia test: a new test of imitation of gestures for elderly people. *Eur J Neurol* 1998; **5:** 571–578.
20. Rothi LJG, Raymer AM, Heilman KM. Limb praxis assessment. In: Rothi LJG, Heilman KM, eds. *Apraxia: The neuropsychology of action*. New York: Psychology Press, 1997: 61–73.
21. Haaland KY, Flaherty D. The different types of limb apraxia errors made by patients with left vs. right hemisphere damage. *Brain Cogn* 1984; **3:** 370–384.
22. Helm-Estabrooks N. *Test of Oral and Limb Apraxia*. Chicago, IL: Riverside Publishing Company, 1992.
23. Kaya K, Unsal-Delialioglu S, Kurt M, Altinok N, Ozel S. Evaluation of ideomotor apraxia in patients with stroke: a study of reliability and validity. *J Rehabil Med* 2006; **38:** 108–112.
24. Poeck K. The clinical examination for motor apraxia. *Neuropsychologia* 1986; **24:** 129–134.
25. Zwinkels A, Geusgens C, van de Sande P, Van Heugten C. Assessment of apraxia: inter-rater reliability of a new apraxia test, association between apraxia and other cognitive deficits and prevalence of apraxia in a rehabilitation setting. *Clin Rehabil* 2004; **18:** 819–827.
26. Alexander MP, Baker E, Naeser MA, Kaplan E, Palumbo C. Neuropsychological and neuroanatomical dimensions of ideomotor apraxia. *Brain* 1992; **1:** 87–107.
27. van Heugten CM, Dekker J, Deelman BG, van Dijk AJ, Stehmann-Saris JC, Kinebanian A. Outcome of strategy training in stroke patients with apraxia: a phase II study. *Clin Rehabil* 1998; **12:** 294–303.
28. Roy EA, Heath M, Westwood D, *et al*. Task demands and limb apraxia in stroke. *Brain Cogn* 2000; **44:** 253–279.
29. Goldenberg G, Laimgruber K, Hermsdorfer J. Imitation of gestures by disconnected hemispheres. *Neuropsychologia* 2001; **39:** 1432–1443.
30. Goldenberg G, Strauss S. Hemisphere asymmetries for imitation of novel gestures. *Neurology* 2002; **59:** 893–897.
31. Fleiss J. Reliability of measurement. In: *The Design and Analysis of Clinical Experiments*. New York: John Wiley and Sons, 1986: 1–31.
32. Landis JR, Koch GG. An application of hierarchical kappa-type statistics in the assessment of majority agreement among multiple observers. *Biometrics* 1977; **33:** 363–374.
33. Hanna-Pladdy B, Heilman KM, Foundas AL. Cortical and subcortical contributions to ideomotor apraxia: analysis of task demands and error types. *Brain* 2001; **124:** 2513–2527.
34. Leiguarda R. Limb apraxia: cortical or subcortical. *Neuroimage* 2001; **14:** S137–S141.
35. Hanna-Pladdy B, Daniels SK, Fieselman MA, *et al*. Praxis lateralization: errors in right and left hemisphere stroke. *Cortex* 2001; **37:** 219–230.
36. Carmo JC, Rumiati RI. Imitation of transitive and intransitive actions in healthy individuals. *Brain Cogn* 2009; **69:** 460–464.
37. Cubelli R, Bartolo A, Nichelli P, Della Sala S. List effect in apraxia assessment. *Neurosci Lett* 2006; **407:** 118–120.
38. Goldenberg G. Defective imitation of gestures in patients with damage in the left or right hemispheres. *J Neurol Neurosurg Psychiatry* 1996; **61:** 176–180.
39. Van Heugten CM, Dekker J, Deelman BG, Stehmann-Saris FC, Kinebanian A. A diagnostic test for apraxia in stroke patients: internal consistency and diagnostic value. *Clin Neuropsychol* 1999; **13:** 182–192.

Appendix 1 – TULIA items*

Imitation, non-symbolic

1. Put index finger on top of nose
2. Bring thumb extended on forehead, other fingers point upwards
3. Bring back of the hand under chin, shoulder 90° abducted
4. Place the hand flat on top of the head
5. Lift only the hand from the table (forearm stays on the table)
6. Spread little finger outwards
7. Extend the arm sideward up to shoulder height
8. Lift middle finger

Imitation, intransitive

9. Make a catholic cross sign
10. Show as if someone is crazy[§]
11. Wipe dust from shoulder[§]
12. Salute like a soldier
13. Hitch for a car[§]
14. Make a stop sign
15. Clasp fingers
16. Point to a bird in the sky

Imitation, transitive

17. Drink from a glass
18. Comb hair[§]
19. Pick up telephone

20. Smoke a cigarette
21. Use a hammer§
22. Use a key
23. Use scissors§
24. Use a stamp to post-mark

Pantomime, non-symbolic

25. *Place your hand flat on your head*
26. Put your hand on your right (or left) shoulder
27. Take your left (or right) ear between thumb and index finger
28. *Put your extended thumb on your forehead, other fingers point upwards*
29. *Extend your arm sideward up to shoulder height*
30. Bend your elbow and look at the palm of your hand
31. *Lift only your hand from the table*
32. Lift your index finger from the table

Pantomime, intransitive

33. *Salute like a soldier*
34. Throw me a kiss
35. *Show as if someone is crazy*§
36. Scratch your head§
37. *Point to a bird in the sky*
38. Wave goodbye§
39. *Make a stop sign*
40. Make a threatening sign

Pantomime, transitive

41. Brush your teeth§
42. *Comb your hair*§
43. Eat soup§
44. *Smoke a cigarette*
45. Use a screwdriver§
46. *Use a key*
47. *Use a stamp to post-mark*
48. Cut bread that is put on the table§

*Within each subtest the first group of four items are proximal, the second group are distal. All non-symbolic items are simple in nature. For symbolic gestures (intransitive and transitive) simple and §repetitive items were selected. 12 items (italics) of the imitation part were repeated in the pantomime domain allowing direct comparison of performances at individual item level.

Appendix 2 – Scoring method

5: Normal movement or identical to the demonstrated movement.

4: Goal of the movement is achieved, but errors occur not affecting trajectory (normal movement plane relative to goal object [tool or own body], normal joint coordination and movement shape).

Movement is too slow, hesitating, robot-like, sloppy with minor spatial errors such as reduced amplitudes.

3: Goal of the movement is achieved, errors subtly affecting trajectory occur, but are corrected.

Extra movements and omissions are present (mainly distal), even brief content errors (substitutions, perseverations) may occur; however, corrections are made in the ongoing movement.

2: Goal of the movement is achieved, errors subtly affecting trajectory occur, but are not corrected.

Body-part-as-object errors, extra movements and omissions (mainly distal) occur without correction.

1: Goal of the movement is not achieved, errors grossly affecting trajectory occur or semantic content is incorrect.

Final position is false, major errors in spatial orientation, overshoot and extra movements (particularly proximal), however, overall movement pattern remains recognizable. Persisting substitutions (related or unrelated) and perseverations occur.

0: No movement, unrecognizable movement.

Seeking and amorphous movements, no temporal or spatial relationship to the requested gesture.

