European Journal of Physical and Rehabilitation Medicine 2025 June;61(3):425-36
DOI: 10.23736/S1973-9087.25.08963-4

ORIGINAL ARTICLE

Psychometric properties of the German Stroke and Aphasia Quality of Life Scale 39 generic version

Caterina BREITENSTEIN [1] *, Katerina HILARI [2], Annette BAUMGAERTNER [3], Tanja GREWE [4], Agnes FLÖEL [5], Wolfram ZIEGLER [6], Peter MARTUS [7], Erich B. RINGELSTEIN [1], Walter HUBER [8], Karl G. HAEUSLER [9], Stefanie BRUEHL [8, 10, 11], Klaus WILLMES [12], on behalf of FCET2EC Study Group ‡

[1]Department of Neurology, University of Muenster, Muenster, Germany; [2]School of Health and Medical Sciences, City St George's, University of London, London, UK; [3]Department of Speech and Language Therapy, Institute of Health Sciences, University of Luebeck, Luebeck, Germany; [4]Department of Technology and Health for People, Jade University of Applied Sciences, Oldenburg, Germany; [5]Department of Neurology, University Medicine of Greifswald, Greifswald, Germany; [6]Clinical Neuropsychology Research Group (EKN), Institute of Phonetics and Speech Processing, Ludwig-Maximilians-University, Munich, Germany; [7]Institute for Clinical Epidemiology and Applied Biometry, University of Tuebingen, Tuebingen, Germany; [8]Clinical and Cognitive Neuroscience, Department of Neurology, RWTH Aachen University, Aachen, Germany; [9]Department of Neurology, University Hospital Ulm, Ulm, Germany; [10]St. Mauritius Rehabilitation Centre, Meerbusch, Germany; [11]Medical Faculty, Heinrich-Heine University Duesseldorf, Duesseldorf, Germany; [12]Department of Neurology, RWTH Aachen University, Aachen, Germany

‡Members are listed at the end of the paper

*Corresponding author: Caterina Breitenstein, Department of Neurology, University Hospital Muenster, Albert-Schweitzer-Campus 1, bldg. A1, 48149 Muenster, Germany, E-mail: breitens@uni-muenster.de

ABSTRACT

BACKGROUND: The international expert consensus core outcome set for post-stroke aphasia recommends the Stroke and Aphasia Quality of Life Scale - 39/generic (SAQOL-39g) for assessing patient-reported health-related quality of life. Cultural adaptations of the SAQOL-39g are mandatory in stroke rehabilitation.
AIM: We adapted the original English SAQOL-39g into German and evaluated its psychometric quality.
DESIGN: Evaluation of a self-report scale embedded in a prospective multicenter parallel group randomized waitlist-controlled trial on the effectiveness of intensive speech and language therapy.
SETTING: Nineteen in- and outpatient aphasia rehabilitation centers in Germany.
POPULATION: People with chronic post-stroke aphasia (N.=156) of all types and severity levels.
METHODS: We followed applicable guidelines for cross-cultural test adaptations and psychometric evaluations. Psychometric analyses are based on the assessment before three weeks of intensive speech and language therapy (acceptability, internal consistency, validity; N.=156), on the assessments before and after three weeks of waiting in the control group (test-retest reliability; N.=78), and on the assessments before and after three weeks of intensive speech and language therapy (responsiveness; N.=156).
RESULTS: The German SAQOL-39g was feasible across all aphasia severity grades (no missing data; no floor/ceiling effects). Internal consistency was excellent (Cronbach's α=0.90); test-retest reliability was moderate-to-good (intraclass-correlations: ICC=0.73 for single/0.85 for average measures). Both exploratory factor analyses and multidimensional scaling of proximity data/graphical network analysis supported the 3-dimensional structure (domains: physical, psychosocial, communication) of the English original version. Convergent (|r|=0.29 to 0.48) and discriminative (|r|=0.03 to 0.07) validities were acceptable. Responsiveness to intervention-induced change showed a small-to-medium treatment effect (group difference after intervention compared to waiting-list control: Cohen's d=0.34).
CONCLUSIONS: The German SAQOL-39g is a reliable, valid and change-sensitive patient-reported outcome measure to assess the physical, communication and psychosocial quality of life in chronic post-stroke aphasia, with comparable psychometric properties and factorial structure to the original English version.

CLINICAL REHABILITATION IMPACT: The German SAQOL-39g is an easy-to-administer and -score patient-reported scale that can be used in rehabilitation settings to measure health-related quality of life and support patient-centered goal setting in people with chronic post-stroke aphasia of different ages, stroke durations, severity and type of aphasia.

(*Cite this article as:* Breitenstein C, Hilari K, Baumgaertner A, Grewe T, Flöel A, Ziegler W, *et al.;* FCET2EC Study Group. Psychometric properties of the German Stroke and Aphasia Quality of Life Scale 39 generic version. Eur J Phys Rehabil Med 2025;61:425-36. DOI: 10.23736/S1973-9087.25.08963-4)

KEY WORDS: Stroke rehabilitation; Aphasia; Quality of life; Psychometrics.

Since introduction of the International Classification of Functioning, Disability and Health (ICF) by the Word Health Organisation,[1] the standardized assessment of patient-reported outcome measures (PROMs) is considered a key priority in post-stroke rehabilitation research. Of particular importance is the assessment of health-related quality of life (HRQoL), *i.e.*, the effects of the disease on the physical, psychological and social quality of life. People with aphasia (PWA) require easily accessible versions of all patient-reported scales because of their language impairments. One of the most widely used patient-reported measures to assess HRQoL in post-stroke aphasia[2] is the Stroke and Aphasia Quality of Life Scale - 39 item version (SAQOL-39[3, 4]). Its 39 questions, presented in a standardized interview, refer to self-perceived functioning in daily activities or feelings during the previous week, each rated on a scale ranging from one (poor HRQoL) to five (excellent HRQoL).

The original English version of the SAQOL-39 was co-developed with PWA in the United Kingdom by adapting the items of the Stroke Specific Quality of Life Scale (SS-QOL).[5] A first psychometric evaluation in a sample of N.=83 PWA in the chronic stage post-stroke with varying degrees of verbal production deficits (yet relatively preserved language comprehension) yielded an overall summary score and a 4-dimensional structure with the domains physical (17 items), psychosocial (11 items), communication (7 items) and energy (4 items) related HRQoL.[4] A subsequent evaluation in a generic stroke sample included stroke survivors with and without aphasia assessed six months post-stroke (N.=71) favored a 3-dimensional solution (domains: physical, psychosocial, communication) for the same pool of 39 questions.[3] All four items of the SAQOL-39 energy domain and one item of the physical domain (item SR7: effect of physical problems on social life) grouped with the SAQOL-39g psychosocial domain. The remaining 34 items grouped on the same domain in both versions.

Both the 3- and the 4-dimensional English versions demonstrate good psychometric properties. Preference is given to the 3-dimensional SAQOL-39g as it can be administered to stroke patients with and without aphasia, beginning with the late subacute stage after stroke, thus allowing for comparisons among different stroke subgroups. Multiple language adaptations of the SAQOL-39/SAQOL-39g exist (Supplementary Digital Material 1: Supplementary Table I),[6, 7] with varying psychometric quality.[8, 9] A German adaptation is not yet available even though German is the most widely spoken native language within the European Union.

The SAQOL-39/SAQOL-39g has been recommended with a 96% consensus by the international Research Outcome Measurement in Aphasia/ROMA consensus statement[10] as the core outcome set (COS) measure for quality of life in all aphasia treatment studies. There is thus a pressing need for language adaptations of the SAQOL-39g where not yet available. We aimed to address this need and present here the psychometric properties of the German SAQOL-39g adaptation, evaluated in a sample of 156 PWA in the chronic stage post-stroke as part of the multicenter randomized controlled trial FCET2EC[11] (pre-registered with ClinicalTrials.gov: NCT01540383). In contrast to prior SAQOL-39/SAQOL-39g psychometric evaluation studies, we 1) followed the quality criteria proposed by the COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) group regarding measurement properties of health status scales;[12-14] and 2) included PWA with all aphasia severity levels except for the most extreme (cf., Methods).

Materials and methods

Participants

Nineteen German in- or out-patient rehabilitation centers recruited 156 PWA (100 male, 56 female; aged 18 to 70 years; enrolled ≥6 months after stroke onset) into a randomized trial on the effectiveness of intensive speech and language therapy (SLT). We included PWA of all aphasia severity levels except for the most severe cases (communi-

cation score <1 on the Aachen Aphasia Test [AAT][15] spontaneous speech communication scale, *i.e.*, "yes" or "no" responses were impossible; or <1/10 correct responses on the easiest part 1 of the 50-items AAT Token Test[15]). Demographic, stroke- and aphasia related participant characteristics of the entire sample are reported in Table I. Sample characteristics of the two groups separately (intervention, control), inclusion/exclusion criteria, the study's design, details of the intensive SLT and effectiveness results for improving functional communication have been

TABLE I.—*Participant characteristics (as assessed before intensive SLT).*

Variable		Entire sample (N.=156)
Demographics		
Age in years	M±SD, range	53.19±9.59, 23-70
Sex	n: female/male (percentages)	56/100 (36%/64%)
Education years	median; min.-max.; Q1-Q3	10; 8-19; 10-18
Stroke		
Stroke severity at study onset - mRS, Range: 0-6	median; min.-max.; Q1-Q3	2; 1-4; 2-3
Months post index stroke	median; min.-max.; Q1-Q3	31; 6-235; 14.25-62
Stroke subtype (n)	ischemic	101 (65%)
	ischemic with hemorrhagic transformation	30 (19%)
	hemorrhagic	17 (11%)
	subarachnoid hemorrhage	8 (5%)
Aphasia		
Type (based on AAT syndrome classification)	Global [n (percentage of sample)]	33 (21%)
	Wernicke [n (percentage of sample)]	25 (16%)
	Broca [n (percentage of sample)]	47 (30%)
	Anomic [n (percentage of sample)]	38 (24%)
	Not classifiable [n (percentage of sample)]	13 (8%)
AAT subtest naming (N.=153)	T-score (M±SD)	52.4±7.3
AAT subtest repetition (N.=151)	T-score (M±SD)	51.9±6.3
AAT subtest written language (N.=154)	T-score (M±SD)	50.7±7.7
AAT subtest language comprehension (N.=154)	T-score (M±SD)	54.2±8.3
AAT Token Test (N.=155)	T-score (M±SD)	51.0±7.7
AAT Spontaneous Speech (clinician-rated), score range per scale: 0-5		
Communication**	median; min.-max.; Q1-Q3	2; 1-5; 1.25–3
Articulation	median; min.-max.; Q1-Q3	4; 1-5; 3-4
Automated speech	median; min.-max.; Q1-Q3	4; 0-5; 3-5
Semantic structure	median; min.-max.; Q1-Q3	3;0-5; 3-5
Phonematic structure	median; min.-max.; Q1-Q3	4;0-5; 3-4
Syntactic structure	median; min.-max.; Q1-Q3	2; 0-5; 1-4
Aphasia severity, AAT profile height*	T-score (M±SD)	51.6±6.1
Aphasia severity classification (based on AAT profile height T-score*)	Minimal, T-score ≥ 62.5 (n)	6 (4%)
	Mild, T-score 52.5-62.4 (n)	64 (41%)
	Medium, T-score 42.5-52.4 (n)	79 (51%)
	Severe, T-score < 42.5 (n)	7 (5%)
General language ability (SAPS total score, score range: 0 – 900; N.=146)	median	490.63
	min.-max.	102.75-757.25
	Q1-Q3	364.56-450.44
Functional communication		
Direct behavioural observation (ANELT), score range 10-50	median	28.25
	min.-max.; Q1-Q3	10-48.50; 20.50-39.33
Mood		
VAMS, mean score across the 6 negative mood items (score range: 0-100; N.=155)	median	17.71
	min.-max.; Q1-Q3	0-84.70; 6.01-28.61
VAMS, sadness item (score range: 0-100; N.=155)	median	9.80
	min.-max.; Q1-Q3	0-99; 2.94-28.43
Cognition		
General intellectual functioning (WAIS-R Picture Completion; score range: 0-16; N.=156)	median	9
	min.-max.; Q1-Q3	0-16; 5-11.75
Auditory short-term memory (WMS-R verbal span forward; score range: 0-12; N.=153)	median	2
	min.-max.; Q1-Q3	0-10; 0-4

n: sample size; M: mean; SD: standard deviation; mRS: modified Rankin Scale; Q1: first quartile; Q3: third quartile; ANELT: Amsterdam Nijmegen Everyday Language Test; VAMS: Visual Analog Mood Scales; WAIS-R: Wechsler Adult Intelligence Scale – Revised; Wechsler Memory Scale – Revised; AAT: Aachen Aphasie Test; AAT profile height: average weighted T-scores of the AAT subtests; *imputed T-scores with <2 percent of raw data missing; **participants had to score at least 1 on the AAT Spontaneous Speech Scale, Communication rating, as part of the study inclusion criteria.

published before.[11] All participants (and if required their legal representative) had given written informed consent for study participation prior to inclusion.

Procedure and measures

The SAQOL 39g has been adapted from the SS-QOL[5] through consultation with PWA and professionals in the UK;[16] full details in.[17] The scale is an interviewer-facilitated self-report scale comprising 39 questions across three HRQoL domains (physical: 16 items; communication: 7 items; psychosocial: 16 items). Twenty-one items refer to everyday life activities (*e.g.*, "How much trouble did you have getting dressed/speaking…?"); the remaining 18 items refer to feelings (*e.g.*, "Did you feel withdrawn from other people?") and other appraisals (*e.g.*, "Did you go out less often than you would like?"). The item response format varies from 1 (very low HRQoL) to 5 (unaffected HRQoL). The timeframe for all questions is the previous week. Item presentation is multimodal (oral/written); responses can be provided verbally or nonverbally (pointing to a written option, gesturing), supported by the interviewer. The SAQOL-39g yields a total and three domain scores ranging from 1 to 5, respectively, calculated by summing up the item scores and dividing by the number of items. The interviewer needs to have skills in supported communication with PWA and follow administration guidelines available online (https://cityaccess.org/tests/saqol-39g). Scoring requires no training.

For the German SAQOL-39g adaptation, we followed established guidelines for cross-cultural adaptations of self-report measures[18, 19] including initial translation (3 translators of whom 2 were bilingual English-German speakers and were aphasia experts), synthesis of the translated versions through discussion between translators, back-translation (by 2 bilingual German-English translators of whom one was naïve regarding aphasia and measuring HRQoL), and expert committee review, resulting in appraisal of the final German version by the original developer K.H. and a feedback report (see Supplementary Digital Material 2: Supplementary Text File 1 for the consensus version of the German-SAQOL-39g). All translators worked independently of each other and from the original developer. Comprehensiveness and comprehensibility of the German version were ensured in a pilot study with n=10 PWA in the chronic stage post-stroke (age: 34–67 years; 3 females; mean time after stroke: 6.3±3.4 years, Broca's aphasia: N.=7, Wernicke's aphasia: N.=1, global aphasia: N.=2; severe to moderate aphasia severity indicated by a mean AAT profile T-score of 46.9±2.6), administered by interviewers skilled in supported communication and familiar with the administration guidelines of the English original version.

The German adaptation of the SAQOL-39g was psychometrically evaluated as part of the FCET2EC trial.[11] The study was approved by the institutional research ethics committee of the lead trial physician (A.F.) at the Charité - Universitaetsmedizin Berlin, Berlin, Germany (protocol number: EA1/234/11; chairperson during study conduct: Prof R. Uebelhack, MD; date of initial approval: 8th Dec 2011) before the trial started. Study implementation was in line with the principles set forth in the 2024 World Medical Association Declaration of Helsinki.[20] The study design comprised two parallel groups (Figure 1). Half of the PWA (N.=78) were randomly assigned to immediate SLT (10 h/week with therapist plus 5 h/week self-managed exercises for at least 3 weeks); the other half (N.=78) was randomly assigned to a waiting-list control group (3 weeks of waiting with usual care) prior to receiving the same intensive SLT regimen as the intervention group. Psychometric evaluations were based on the baseline assessment immediately prior to SLT (T2) unless otherwise indicated (see test-retest reliability and responsiveness analyses below).

Data analysis

Composite scores (total, 3 domains) were calculated as arithmetic mean scores across items, as for the original English version, and were based on the number of valid items, respectively. Imputation of missing data was not required (total missing data: 3 item responses of which 2 occurred in a single participant).

For psychometric evaluation, we applied the quality

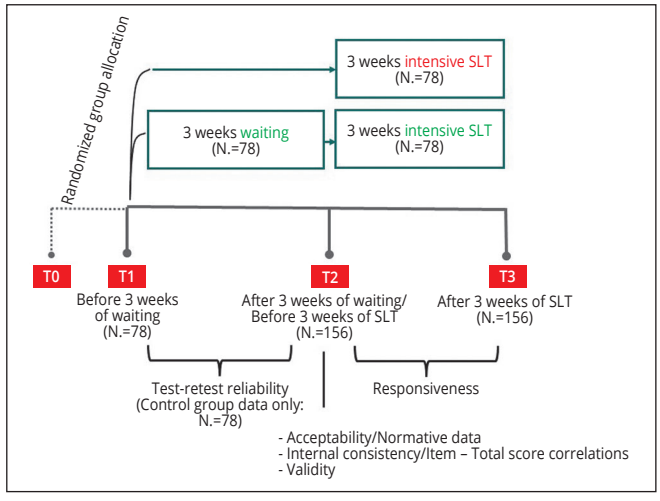

Figure 1.—FCET2EC study design and assessments used for psychometric analyses.

criteria framework proposed by the COSMIN group[12, 14] for patient-reported outcomes. For comparability with the English SAQOL-39g, we applied the same a priori defined psychometric benchmarks (as per Hilari *et al.*, 2009[3]).

1) Objectivity:

• administration and scoring followed standardized procedures,[3] such as providing clear and written instructions for assessors and establishing specific criteria for scoring responses;

• normative data for score interpretation were developed based on the approach reported in[21] for calculating percentile ranks (PRs). The online Psychometrica norm score calculator (https://www.psychometrica.de/normwertrechner_en.html) was used to additionally convert ordinal-scaled PRs into interval-scaled T-scores.

2) Acceptability:

• ≤10% of missing data and ≤80% of floor/ceiling effects for a given item

• the ratio of skewness and kurtosis to their standard error, respectively, is in absolute terms ≤3.29[22] for at least 75 percent of the items to assume approximately normal distribution of scores.

3) Reliability:

• internal consistency: Cronbach's α>0.70 (total scale and domains) and its calculation based on ≥100 participants; McDonald omega (ω) was calculated in addition to Cronbach's alpha to account for the multidimensionality of the SAQOL-39g (benchmark: ω>0.70);

• item total correlations ≥0.30 (Pearson correlation coefficients);

• test-retest reliability: Test-retest reliability was determined using the data before and after the waiting period in the control group (Figure 1: T1 *versus* T2; N.=78). Following COSMIN recommendations,[23] the benchmark was set at Intraclass Correlation Coefficients/ICCs ≥0.70 for the appropriate ICC model (two-way mixed model, absolute agreement and single measure). For ease of comparison with the English[3] and other language adaptation studies,[8] we additionally report results for ICCs based on a two-way mixed model, assessing consistency and average measure.

4) Validity:

• Content validity: Content validity was tested for the English SAQOL-39 in the United Kingdom.[16, 17] Given the geographical proximity and the close cultural similarities between Germany and the UK, and the rigorous linguistic adaptation process, we did not expect differences with respect to the relevance, comprehensiveness and comprehensibility of the items in our German stroke sample. These assumptions had been confirmed in a pilot study with N.=10 German-speaking PWA in the chronic stage post-stroke.[24]

• Internal validity: moderate total and domain score intercorrelations; moderate correlations between domains. We report Pearson correlation coefficients.

• Structural validity: COSMIN recommends using exploratory factor analysis (principal component analysis/PCA and maximum likelihood factor analysis [MLFA]) using a sample size of at least five times the number of items, *i.e.*, n≥39×5=195. In this study, sample size was close to the required criterion (N.=156). MFLA rather than Principal Axis Factoring (PAF) was chosen as MLFA outperformed PAF when the number of selected factors is based on theoretical considerations.[25] MLFA yielding a 3-factor solution was conducted with varimax and promax rotations, respectively. Promax (oblique) rotation with Kaiser normalization was included because all HRQoL domains may be intercorrelated to a certain degree. The threshold for Eigenvalue was set to ≥1; factor loading to >0.40; cross-loading was defined as a difference <0.20 between loadings on more than one factor (with a factor loading of >0.40 on at least one of these factors). We also applied exploratory nonmetric/ordinal multidimensional scaling (PROXSCAL) to visualize the representation of items in a two-dimensional space and graphical network analysis (using jasp.org).

• Construct validity: For convergent validity, SAQOL-39g scores at T2 were correlated with measures a priori assumed to be related to the three SAQOL-39g subdomains: motor functioning (modified Rankin Scale/mRS;[26]), general language (total score of the SAPS-'Sprachsystematisches Aphasiescreening'),[27, 28] reading and writing performance (AAT subtest Written language),[15] functional communication (Amsterdam Nijmegen Everyday Language Test/ANELT; A-scale),[29] and emotional well-being (Visual Analog Mood Scales/VAMS).[30] For discriminative validity, correlation coefficients were calculated between SAQOL-39g scores (at T2) and measures a priori assumed to be unrelated to HRQoL. These were general intellectual functioning (subtest Picture Completion of the German adaptation of the Wechsler Adult Intelligence Scale);[31] as well as auditory short-term memory (subtest "digit span forward" of the German adaptation of the Wechsler Memory Scale).[32] Reported P-values for correlation coefficients (Pearson and Spearman rank) were uncorrected for multiple comparisons because of the strong prior evidence (based on the original English version) on which correlations were likely to be significant. Acceptable construct validity was assumed if ≥75 percent of results were in accordance with prespecified hypotheses (moderate to high correlations for similar and

related constructs: |r|≥0.30, and low correlations for unrelated constructs: |r|<0.30).

• Cross-cultural validity: At the time this study was planned, cross-cultural validity had not been included yet in the COSMIN psychometric grading framework and we had not planned to perform a multi-group confirmatory factor analysis (MGCFA) to compare the English and the German samples. MGCFA requires at least N.=150 participants per group (or five to seven times the number of items) according to COSMIN. Because the English aphasia sample (N.=83 participants) did not match the required sample size, MGCFA was not pursued. Comparisons between the German and English aphasia samples and psychometric properties are therefore purely descriptive (Supplementary Table I).[3, 4] Within the German sample, we calculated means and standard deviations for different ages (2 levels: working age *versus* >65 years) and aphasia severity (2 levels based on the AAT profile score: T-score ≥52.5 "minimal/mild" *versus* T-score <52.5 "moderate/severe"). Groups were compared using independent samples *t*-tests (with P≤0.05 as significant, two-sided).

5) Responsiveness:

• Responsiveness was analyzed by comparing SAQOL-39g scores (i) before and after three weeks of intensive SLT (T2 *versus* T3) using paired *t*-tests and (ii) between groups after three weeks of SLT (intervention group at T3) *versus* three weeks of waiting (control group at T2) with baseline performance (T2 for intervention group; T1 for control group) as covariate using ANCOVA with P≤0.05 as significant. Intervention responsiveness data for the current sample have been published before,[11] but domain scores in that publication were based on the 4-factors structure of the SAQOL-39.[4]

• Treatment effect sizes are reported as Cohen's *d* based on the *F*-value for the group difference after three weeks of SLT (intervention group: N.=78) *versus* three weeks of waiting (control group: N.=78) with baseline performance as covariate. Additionally, we analyzed treatment effect size estimates pooled across groups for repeated measures with pooled standard deviations (N.=156; pre-post SLT: $d_{RM,pooled}$) which take the correlation between repeated assessments into account and standardized response means (SRMs; for repeated measures from before to after 3 weeks of SLT). For SRM, the mean score change (T3 minus T2) was divided by the SD of the change score.

• To provide benchmarks for *individual* score changes, we also calculated the Smallest Detectable Change/SDC (smallest statistically significant change score for an individual) and a Minimal Important change/MIC (the smallest change score from pre to post intervention considered clinically relevant by relevant stakeholders). The SDC for the German evaluation sample has been reported before.[33, 34] The FCET2EC trial had recruited participants from 2012-2015 and lacked a patient-reported "anchor" measure of treatment success from the patients' perspective, as recommended in more recent publications.[34, 35] For calculation of a MIC, we therefore followed the anchor-based approach outlined by,[7] in that an improvement of at least one level on the modified Rankin Scale/mRS from before to after intervention is a clinically meaningful difference in post-stroke aphasia. We report the MIC as the mean SAQOL-39g total score change from before to after intervention for participants who improved at least one level on the mRS between these two assessments (treatment "responders"). All other participants were classified as non-responders. We used the "mean change method" instead of a predictive modelling approach[36] to determine the MIC benchmark because of the unequal distribution of treatment responders (14%) and non-responders (86%) on the mRS[37] and because the COSMIN criterion of a minimum sample size of N.=30 per (responder) group was not met.[14] Mean total change scores from before to after the intensive SLT intervention for treatment responders (N.=22) and non-responders were compared using Mann-Whitney-U-tests; mean change scores from before to after the intensive SLT intervention within the responder/non-responder groups were compared using Wilcoxon signed rank tests, respectively.

• We also applied the "criterion approach" suggested by COSMIN[14] and examined whether a criterion correlated substantially with the SAQOL-39g change scores from before to after the SLT intervention using Pearson's and Spearman's rank correlation coefficients. In the absence of an appropriate criterion for HRQoL, we used mRS change scores as in previous research.[7]

Data availability

Part of the data (N.=142/156 participants consented to anonymous data sharing) associated with the paper are available on request from the "data sets" repository of the Collaboration of Aphasia Trialists (CATs; https://www.aphasiatrials.org/aphasia-dataset; last accessed on 13th January 2025).

Results

Supplementary Table I details the psychometric properties of the German SAQOL-39g and contrasts them with the two English versions.

1) Objectivity:

• Assessors used the written instructions which had been developed for the English version. Training was provided for administration of the scale in a 1-day in-person workshop led by the first author (C.B.). As part of the training procedure, each assessor also completed the scale with a sample patient. The FCET2EC study centre provided individual feedback regarding correct completion and documentation of the SAQOL-39's response sheet.

• Normative data (PRs and T-scores) for the total scores of the German SAQOL-39g based on the current sample of N.=156 PWA (≥6 months post-stroke) are provided in Supplementary Digital Material 3: Supplementary Table II.

2) Acceptability:

• Acceptability of the German SAQOL-39g was high in all 156 participants with chronic post-stroke aphasia (no missing data and no floor/ceiling effects as per a priori defined criteria).

• In terms of score distributions, skewness and kurtosis values indicated significant departure from symmetry and peakedness relative to a normal distribution for >25 percent of the items (skewness: 24/39 items =61.5%; kurtosis: 10/39 items =25.6%). Negative skewness (*i.e.*, higher QoL scores) was prominent for physical/psychosocial domain items, whereas language domain items presented with positive skewness (*i.e.* lower QoL scores), a pattern to be expected in a sample of chronic post-stroke aphasia. Kurtosis was positive (a more peaked distribution than normal) for physical domain items and negative (a flatter distribution than normal) for psychosocial domain items.

3) Reliability:

• internal consistency was excellent (Cronbach's α and McDonald ω: total score, α/ω=0.90; domain scores: α=0.80-0.91/ω=0.79-0.91);

• item total correlations were all $r \geq 0.30$ as required;

• test-retest reliability was moderate-to-good (total score: ICC=0.73 for single/0.85 for average measures; domains: ICC=0.64-0.84 for single/0.78-0.91 for average measures). The ICC for the German SAQOL-39g (N.=78) was substantially lower than for the English version (based on N.=18, see Supplementary Table I). For comparability with the English SAQOL-39g evaluation sample, we analyzed ICCs for a subgroup of *n*=53/78 with only moderate-mild auditory comprehension impairments (AAT Token Test T-score ≥ 50). As expected, test-retest reliability increased with a more homogenous and less severe aphasia sample (total score: ICC=0.80 for single measure/0.89 for average measure; domains: ICC=0.67-0.89 for single measure/0.81-0.95 for average measure).

4) Validity:

• internal validity was good, with moderate intercorrelations between total and domain scores ($0.65 \leq r \leq 0.85$) and moderate domain score intercorrelations ($0.24 \leq r \leq 0.53$).

• structural validity assessment using PCA and MLFA (Supplementary Table I for overall results and Supplementary Digital Material 4: Supplementary Table III for factor loadings) supported the three-factors solution of the original SAQOL-39g and showed similar factor loadings of the items, with only 3/39 items showing maximum loading on a different factor compared to the English version (T5, FR9, SR8), four items crossloading on two factors (MD3/7, FR7/9), and five items not loading >0.40 on any of the three factors. Exclusion of these cross- or low-loading items did not improve the scale's internal consistency (Supplementary Digital Material 5: Supplementary Table IV). Multidimensional scaling using PROXSCAL and graphical network analysis also supported the three-factors solution (Supplementary Digital Material 6: Supplementary Figure 1). The sample size of *N.*=156 was close to the COSMIN criteria of N.≥195.

• Construct validity was good and in the expected range,[14] both for convergent (Pearson: total score: $|r|$=0.29 to 0.48); domains: $|r|$=0.30-0.63) and discriminative (Pearson: total score: $|r|$=0.03 to 0.07; domains: $|r|$=0.01 to 0.15) validities, with ≥75 percent of the correlations in the expected direction (Supplementary Table I and Supplementary Digital Material 7: Supplementary Table V which also lists results for Spearman rank correlations).

• Cross-cultural validity: The English[4] and German aphasia samples (Table I) were highly similar regarding sex distributions (>60% males), stroke chronicity (on average >2.5 years after the initial stroke) and aphasia severity (about 50% classified as "mild-minimal"). The only difference was that the English sample (61.2±15.5 years) was older and more variable in age than the German sample (M=53.2±9.6 years). Means, standard deviations and subgroup comparisons for different aphasia severity and age levels of the German sample are reported in Supplementary Digital Material 8, Supplementary Table VI. There were no age (working-aged *versus* >65 years) effects on SAQOL-39g total or domain scores, but as would be expected, participants with "moderate-severe" as compared to "mild-minimal" aphasia reported lower total as well as lower physical and communication domain scores.

5) Responsiveness:

• Responsiveness comparing SAQOL-39g scores

before and after the intensive intervention yielded significant improvements for the total and all three domain scores (N.=156, all P<0.01; Table II). Total and communication domain scores in the waiting-list control group also significantly increased from before to after the waiting period with usual care (N.=78, P<0.02; Table II). However, improvements from baseline were significantly greater after the intensive SLT intervention than after the waiting period for the total and psychosocial domains (both P<0.04) with a trend towards significant improvement for the communication domain (P=0.07), but no improvement for the physical domain (P=0.34).

• Overall, treatment effect sizes were classified as small (total score, communication and psychosocial domain scores: 0.30≤Cohen's d≤0.35, Table II). Effect sizes for pre-post intervention comparisons pooled across groups were similar ($0.23 \leq d_{Repeated\ Measures,pooled} \leq 0.54$; 0.22≤SRM≤0.42; Supplementary Digital Material 9: Supplementary Table VII).

• For our sample of PWA in the chronic stage post-stroke, the benchmarks for individual total score changes were SDC=0.39 points[33, 34] and MIC=0.24 points (Supplementary Table I and Supplementary Digital Material 10: Supplementary Table VIII). Both mRS responder and non-responder groups improved significantly from before to after the intervention in total SAQOL-39g change scores (*MdN.*=0.24 *versus* 0.15; both P≤0.04); there was no statistically significant group difference (U=1325, P=0.45).

• Criterion approach: The correlation between changes in mRS scores from before to after the intervention and changes in SAQOL-39g total scores was not significant (N.=156; Pearson's r=-0.007, P=0.94; Spearman r_s=0.03, P=0.75).

Discussion

We evaluated the psychometric properties of the German adaptation of the SAQOL-39g, a key measure of the COS for aphasia trials,[10, 38] based on N.=156 PWA in the chronic stage post-stroke. The results support the acceptability, objectivity, reliability, validity and responsiveness of the German SAQOL-39g in this stakeholder group. The German SAQOL-39g psychometric properties were overall highly similar to the English version[3] despite sample differences regarding time post stroke (late subacute stage in the English sample), range of language comprehension impairments (exclusion of very severe/severe receptive aphasia cases in the English sample), and study design (part of a treatment effectiveness trial design for the German evaluation[11]).

Results for some psychometric properties differed from expectation. Skewness and kurtosis values deviated from a normal distribution for more than the a priori defined benchmark of 25 percent of the items. Such a tendency for more extreme scores may be typical for stroke symptoms, which either have a major (scores skewed to the right) or

TABLE II.—*SAQOL-39g scores immediately before and after three weeks of intensive SLT for the entire sample, before and after the waiting period for test-retest assessment (subsample of N.=78 only) and differences between the intervention and control groups after three weeks of intervention* versus *waiting (with baseline as covariate in the ANCOVA).*

		Entire sample (N.=156)			Sub-sample for assessment of test-retest reliability (N.=78)			Group difference (3 weeks intervention *versus* 3 weeks waiting, with baseline as covariate)		
		Pre 3 weeks SLT	Post 3 weeks SLT		pre 3 weeks waiting	post 3 weeks waiting		ANCOVA		Effect size for group difference
		M±SD	M±SD	Paired *t*-test (P value)	M±SD	M±SD	Paired *t*-test (P value)	F(1,153)	P value	Cohen's d (based on F value)
SAQOL-39 g (item score range: 1-5)	Mean total score (39 items)	3.69±0.57 (N.=156)	3.83±0.57 (N.=156)	<0.0001	3.58±0.61 (N.=78)	3.70±0.61 (N.=78)	0.0195	4.54	0.0348	0.34
	Mean physical score (16 items)	4.06±0.70 (N.=156)	4.15± 0.69 (N.=156)	0.0065	3.93±0.79 (N.=78)	4.02±0.75 (N.=78)	0.07	0.91	0.34	0.15
	Mean communication score (7 items)	2.84±0.76 (N.=156)	3.10±0.79 (N.=156)	<0.0001	2.66±0.76 (N.=78)	2.90±0.78 (N.=78)	0.0010	3.42	0.07	0.30
	Mean psychosocial score (16 items)	3.68±0.73 (N.=156)	3.83±0.74 (N.=156)	0.0021	3.63±0.77 (N.=78)	3.72±0.74 (N.=78)	0.20	4.63	0.0330	0.35

N.: sample size; *M:* mean; *SD:* standard deviation; SLT: speech and language therapy; df: degrees of freedom; SAQOL-39: Stroke and Aphasia Quality of Life Scale-39, The effect size Cohen's *d* refers to the average group difference (and standard deviations) of the differences from pre to post assessments.

minor (scores skewed to the left) impact on the person's HRQoL. The observed skewness pattern to the right, with predominantly low HRQoL scores for the communication domain items, is to be expected in a chronic post-stroke aphasia sample including severe cases and thus does not indicate a methodological flaw of the instrument.

Compared to prior SAQOL-39 evaluation studies,[8, 9] test-retest reliability was lower for the German adaptation ('moderate' to 'good' instead of "excellent"[39]), which may be attributable to differences in sample characteristics in comparison with the English version (the latter included stroke survivors with and without aphasia and thus greater between-subject variability resulting in stronger correlation coefficients and excluded PWA with severe language comprehension deficits), retest time interval (7 days in the English sample *versus* 21 days in the German sample), the applied ICC model (which was not explicitly stated in most of the prior evaluation studies) and study design. The intervention study design used here may have triggered treatment expectation effects in some participants of the waiting-list control group (which is one of the reasons for including a waiting-list control group in a clinical trial[40] as indicated by improved mean SAQOL-39g scores from before to after the waiting period). Still, test-retest reliability for the total score (ICC=0.73 for absolute agreement, single measure) was above the benchmark of ICC≥0.70 set by the COSMIN recommendations[14] and is thus acceptable. The 'true' test-retest reliability of the German SAQOL-39g in chronic aphasia may be higher and the issue should be addressed in future evaluation studies not applying an intervention *versus* waitlist control design.

Structural validity of the German SAQOL-39g was adequate in that the 3-factor structure of the original English version was supported by exploratory factor analysis and multi-dimensional scaling based on a sufficiently large sample size.[14] We did not conduct an additional confirmatory factor analysis because we were mainly interested in corroborating the 3-factor solution of the English original version[3] without modifications. Three of 39 items loaded on a different factor in the German as compared to the English version. These items were: 1) T5="Finding it hard to make decisions" (maximum factor loading on the communication instead of the psychosocial domain); and 2) FR9="Language problems effect on family life" and SR8="Language problems effect on social life" (maximum factor loadings on the psychosocial instead of the communication domain, respectively). Item T5 had a maximum factor loading <0.40 in both the German and the English versions. As indicated by the results of the graphical network analysis, this may reflect the ambiguity of T5's item content, such that it can be interpreted literally (expressing decisions through communication) or non-literally (emotional struggling with decisions and thus impacting on psychosocial functioning). For items FR9 and SR8, multidimensional scaling (Supplementary Figure 1) showed that these items were located between the psychosocial and communication item groupings for the German sample. Thus, responses to these items may reflect either the self-perceived communication impairment directly or the psychosocial effects of the communication impairment indirectly. Given the substantial correlation of the psychosocial and communication subdomain scores, this finding is not surprising. Because of the acceptable construct validity of the German version and for cross-cultural comparability with other language versions of the SAQOL-39g, we decided to keep items T5, FR9 and SR8 as in the English version for scoring of the psychosocial (T5) and communication (FR9, SR8) subdomain scores.

Factor loadings represent the strength and direction of the relationship between an item and a particular factor. Traditionally, factor loadings >0.40 have been considered acceptable for inclusion of an item in a scale based on a sample size of ≥200.[41] However, there is no hard-and-fast rule regarding what constitutes an appropriate threshold value for exclusion of an item. In the current evaluation study, 5/39 items (13 percent) were not loading >0.40 on any of the three factors. One item (UE1; "Trouble with writing/typing") loaded to a similar extent (0.28, respectively) on the physical and communication subdomains in the German version, indicating some ambiguity of the item content. For this item, German assessors may not have followed instructions in the manual to refer to the physical challenge of the activity ("*i.e. use your hand to write or type*"). It may be helpful to include explicit instructions directly on the scoring sheet to avoid ambiguity in item interpretation in the future.

The other four items (T4, T5, MD2, MD6) not loading >0.40 on any of the three factors loaded at least 0.31 on the psychosocial factor and had additional factor loadings of >0.19 on the communication factor. If an item has a relatively low factor loading but contributes unique information about the underlying construct, retaining it may help maintain content validity. Supplementary Table III indicates that these five items indeed contribute a high degree of "uniqueness". In the future, larger sample sizes of n>500 will provide more stable estimates of factor load-

ings, making it easier to distinguish meaningful relationships from random noise. For comparability with the English version and other language adaptations, we decided to keep all 39 items in the German version.

In contrast to the English original version and the other language adaptations,[8] our study design allowed to demonstrate treatment responsiveness of the German SAQOL-39g in a chronic aphasia sample including very severe cases. Treatment effects for HRQoL were small-to-moderate and may increase with longer treatment duration than in the current study design. Benchmarks for treatment success on the individual level were provided for minimal statistically significant (SDC; first reported in[33]) as well as minimal clinically meaningful (MIC) score changes. However, MIC calculation was based on the mRS that predominantly assesses the degree of physical independence after stroke. Future studies should base MIC calculation on a more aphasia-relevant "anchor", such as the patient-rated impact of an aphasia intervention.[34] Such meaningful benchmarks are in development,[42, 43] but currently not yet available.

The study was not primarily planned as a psychometric evaluation study for the German SAQOL-39g adaptation, but as proof of efficacy for intensive SLT in chronic aphasia. As a consequence, not all of the strict requirements of the COSMIN framework could be implemented, such as a sample size of N.≥100 participants to determine test-retest reliability or of five times the number of test items (here: N.≥195) for exploratory factor analysis. However, the current overall sample size of N.=156 used for exploratory factor analysis is one of the largest evaluation sample sizes in aphasia research and comes very close to the COSMIN recommendation, thus the results can be considered stable.[44]

We recommend that the psychometric quality criteria be regarded as provisional until results from a methodologically well-planned evaluation study with a larger sample size are available.

Conclusions

In summary, future randomized controlled trials focusing on the efficacy of aphasia-interventions should include the SAQOL-39g for assessing HRQoL from the perspective of PWA as recommended by the international consensus-based COS for aphasia trials.[10, 38] The globally orchestrated administration of the SAQOL-39g as an outcome measure in aphasia trials not only ensures a high psychometric quality in change assessment, but also allows the comparison of study results across cohorts speaking different languages. This will contribute to greater efficiency and higher data quality in aphasia rehabilitation research.[45, 46]

The German adaptation of the SAQOL-39g fits perfectly into these internationalization efforts and represents an accessible, objective, reliable, valid, change-sensitive outcome measure for assessing HRQoL in chronic post-stroke aphasia. It has highly similar psychometric properties to the original English SAQOL-39g and can be recommended for use in both research and clinical settings.

References

1. World Health Organisation. International Classification of Functioning, Disability and Health. Geneva, Switzerland: World Health Organisation; 2001.

2. Neumann S, Quinting J, Rosenkranz A, de Beer C, Jonas K, Stenneken P. Quality of life in adults with neurogenic speech-language-communication difficulties: A systematic review of existing measures. J Commun Disord 2019;79:24–45.

3. Hilari K, Lamping DL, Smith SC, Northcott S, Lamb A, Marshall J. Psychometric properties of the Stroke and Aphasia Quality of Life Scale (SAQOL-39) in a generic stroke population. Clin Rehabil 2009;23:544–57.

4. Hilari K, Byng S, Lamping DL, Smith SC. Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39): evaluation of acceptability, reliability, and validity. Stroke 2003;34:1944–50.

5. Williams LS, Weinberger M, Harris LE, Clark DO, Biller J. Development of a stroke-specific quality of life scale. Stroke 1999;30:1362–9.

6. Enderby P, Wood V, Wade D. Frenchay Aphasia Screening Test. Windsor: NFER-Nelson; 1987

7. Guo YE, Togher L, Power E, Heard R, Luo N, Yap P, *et al.* Sensitivity to change and responsiveness of the Stroke and Aphasia Quality-of-Life Scale (SAQOL) in a Singapore stroke population. Aphasiology 2015;31:427–46.

8. Ahmadi A, Tohidast SA, Mansuri B, Kamali M, Krishnan G. Acceptability, reliability, and validity of the Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39) across languages: a systematic review. Clin Rehabil 2017;31:1201–14.

9. Hilari K. Letter to the Editor re: Ahmadi, A., Tohidast, S. A., Mansuri, B., Kamali, M., & Krishnan, G. Acceptability, reliability, and validity of the Stroke and Aphasia Quality of Life Scale-39 (SAQOL-39) across languages: a systematic review. Clinical Rehabilitation, 2017;31:1201-1214. Clin Rehabil 2020;34:1420–1.

10. Wallace SJ, Worrall L, Rose T, Le Dorze G, Breitenstein C, Hilari K, *et al.* A core outcome set for aphasia treatment research: the ROMA consensus statement. Int J Stroke 2019;14:180–5.

11. Breitenstein C, Grewe T, Flöel A, Ziegler W, Springer L, Martus P, *et al.*; FCET2EC study group. Intensive speech and language therapy in patients with chronic aphasia after stroke: a randomised, open-label, blinded-endpoint, controlled trial in a health-care setting. Lancet 2017;389:1528–38.

12. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, *et al.* Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol 2007;60:34–42.

13. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7.

14. Mokkink LB, Prinsen CA, Patrick DL, Alonso J, Bouter LM, de Vet

HC, *et al.* COSMIN Study Design checklist for Patient-reported outcome measurement instruments [Internet]. ©2019. Available from: https://www.cosmin.nl/wp-content/uploads/COSMIN-study-designing-checklist_final.pdf [cited 2025, Jan 15].

15. Huber W, Poeck K, Willmes K. The Aachen Aphasia Test. Adv Neurol 1984;42:291–303.

16. Hilari K, Byng S. Measuring quality of life in people with aphasia: the Stroke Specific Quality of Life Scale. Int J Lang Commun Disord 2001;36(Suppl):86–91.

17. Hilari K. Assessing Health Related Quality of Life in people with aphasia [Dissertation]. London, UK: City, University of London; 2002.

18. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine 2000;25:3186–91.

19. Mokkink LB, de Vet HC, Prinsen CA, Patrick DL, Alonso J, Bouter LM, *et al.* COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res 2018;27:1171–9.

20. World Medical Association. World Medical Association Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Participants. JAMA 2025;333:71–4.

21. Baumgartner TA. Tutorial: Calculating Percentile Rank and Percentile Norms Using SPSS. Meas Phys Educ Exerc Sci 2009;13:227–33.

22. Kim HY. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 2013;38:52–4.

23. Mokkink LB, Boers M, van der Vleuten CP, Bouter LM, Alonso J, Patrick DL, *et al.* COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Methodol 2020;20:293.

24. Breitenstein C, Korsukewitz C, Baumgärtner A, Flöel A, Zwitserlood P, Dobel C, *et al.* L-dopa does not add to the success of high-intensity language training in aphasia. Restor Neurol Neurosci 2015;33:115–20.

25. de Winter JC, Dodou D. Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. J Appl Stat 2012;39:695–710.

26. van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJ, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. Stroke 1988;19:604–7.

27. Krzok F, Rieger V, Niemann K, Nobis-Bosch R, Radermacher I, Huber W, *et al.* The novel language-systematic aphasia screening SAPS: screening-based therapy in combination with computerised home training. Int J Lang Commun Disord 2018;53:308–23.

28. Bruehl S, Huber W, Longoni F, Schlenck KJ, Willmes K. SAPS – Sprachsystematisches Aphasiescreening [SAPS - Language Systematic Aphasia Screening]. Goettingen: Hogrefe; 2022.

29. Blomert L, Kean ML, Koster C, Schokker J. Amsterdam Nijmegen Everyday Language Test: construction, reliability and validity. Aphasiology 1994;8:381–407.

30. Stern RA. Assessment of mood states in aphasia. Semin Speech Lang 1999;20:33–49, quiz 49–50.

31. Petermann F. Wechsler Adult Intelligence Scale. Fourth Edition. Frankfurt: Pearson; 2012.

32. Lepach AC, Petermann F. Wechsler Memory Scale - Revised. German adaptation. Forth Edition. Bern: Huber; 2012.

33. Menahemi-Falkov M, Breitenstein C, Pierce JE, Hill AJ, O'Halloran R, Rose ML. A systematic review of maintenance following intensive therapy programs in chronic post-stroke aphasia: importance of individual response analysis. Disabil Rehabil 2022;44:5811–26.

34. Breitenstein C, Hilari K, Menahemi-Falkov M, Rose ML, Wallace SJ, Brady MC, *et al.* Operationalising treatment success in aphasia rehabilitation. Aphasiology 2023;37:1693–732.

35. Wallace SJ, Worrall L, Rose T, Le Dorze G. Using the International Classification of Functioning, Disability, and Health to identify outcome domains for a core outcome set for aphasia: a comparison of stakeholder perspectives. Disabil Rehabil 2019;41:564–73.

36. Terwee CB, Peipert JD, Chapman R, Lai JS, Terluin B, Cella D, *et al.* Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. Qual Life Res 2021;30:2729–54.

37. Terluin B, Eekhout I, Terwee CB, de Vet HC. Minimal important change (MIC) based on a predictive modeling approach was more precise than MIC based on ROC analysis. J Clin Epidemiol 2015;68:1388–96.

38. Wallace SJ, Worrall L, Rose TA, Alyahya RS, Babbitt E, Beeke S, *et al.* Measuring communication as a core outcome in aphasia trials: results of the ROMA-2 international core outcome set development meeting. Int J Lang Commun Disord 2023;58:1017–28.

39. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;15:155–63.

40. Hohenschurz-Schmidt D, Vase L, Scott W, Annoni M, Ajayi OK, Barth J, *et al.* Recommendations for the development, implementation, and reporting of control interventions in efficacy and mechanistic trials of physical, psychological, and self-management therapies: the CoPPS Statement. BMJ 2023;381:e072108.

41. Stevens JP. Applied Multivariate Statistics for the Social Sciences. Fifth Edition. London: Routledge; 2012.

42. Harvey S, Stone M, Zingelman S, Copland DA, Kilkenny MF, Godecke E, *et al.* Comprehensive quality assessment for aphasia rehabilitation after stroke: protocol for a multicentre, mixed-methods study. BMJ Open 2024;14:e080532.

43. Zingelman S, Cadilhac DA, Kim J, Stone M, Harvey S, Unsworth C, *et al.* 'A Meaningful Difference, but Not Ultimately the Difference I Would Want': A Mixed-Methods Approach to Explore and Benchmark Clinically Meaningful Changes in Aphasia Recovery. Health Expect 2024;27:e14169.

44. Guadagnoli E, Velicer WF. Relation of sample size to the stability of component patterns. Psychol Bull 1988;103:265–75.

45. Breitenstein C, Wallace SJ, Gilmore N, Finch E, Pettigrove K, Brady MC; with the CATs Executive Committee. Invaluable Benefits of 10 Years of the International Collaboration of Aphasia Trialists (CATs). Stroke 2024;55:1129–35.

46. Stern RA, Arruda JE, Hooper CR, Wolfner GD, Morey CE. Visual analogue mood scales to measure internal mood state in neurologically impaired patients: description and initial validity evidence. Aphasiology 1997;11:59–71.

Conflicts of interest

Caterina Breitenstein received research support from the German Federal Ministry of Education and Research (grant # 01GY1144) and from the German non-profit Society for Aphasia Research and Treatment (GAB) during FCET2EC trial conduct (February 2012 to April 2015). The sponsor was not specifically involved in the research. Katerina Hilari is the developer of the original English SAQOL-39/SAQOL-39g. SB is the developer of the aphasia screening test SAPS; Walter Huber and Klaus Willmes are coauthors of the SAPS (see sections on convergent validity). Agnes Flöel had speaker contracts for Eli Lilly, Biogen Idec, Eisai, and Roche, and advisory board contracts for Eli Lilly and Biogen Idec. Karl G. Haeusler reports consultant relationships with the following companies: Alexion, AstraZeneca, Bayer, Boston Scientific, Daiichi Sankyo, Edwards Lifesciences, Medtronic, Pfizer, Portola, Premier Research. KGH received honoraria for lectures from: Abbott, AstraZeneca, Bayer, Biotronik, Boehringer Ingelheim, Bristol-Myers Squibb, Daiichi Sankyo, Novartis, Pfizer, Sanofi, SUN Pharma, and W.L. Gore and Associates. The remaining authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Funding
This research was funded by grants from the German Federal Ministry of Education and Research (grant # 01GY1144) and the German Society for Aphasia Research and Treatment (GAB). The authors report no involvement in the research by the sponsor that could have influenced the outcome of this work.

Authors' contributions
Caterina Breitenstein and Katerina Hilari shared first authorship based on contribution. All authors except Katerina Hilari were members of the FCET2EC study group. Annette Baumgaertner, Caterina Breitenstein, Agnes Flöel, Wolfram Ziegler, Tanja Grewe, and Peter Martus devised the trial study protocol. E. Bernd Ringelstein, Peter Martus, Walter Huber, Klaus Willmes, and Karl Georg Haeusler were members of the FCET2EC Trial Steering Committee. Caterina Breitenstein, Katerina Hilari, Klaus Willmes, and Peter Martus contributed to data analysis and interpretation. Caterina Breitenstein wrote the first draft of the report with input from Katerina Hilari, Klaus Willmes, Peter Martus, Annette Baumgärtner, Agnes Flöel, Wolfram Ziegler, Tanja Grewe, Karl Georg Haeusler, E. Bernd Ringelstein, Stefanie Bruehl, and Walter Huber. Caterina Breitenstein edited the report. All authors read and approved the final version of the manuscript.

Group author members
Caterina BREITENSTEIN; Annette BAUMGAERTNER; Tanja GREWE; Agnes FLÖEL; Wolfram ZIEGLER; Peter MARTUS; Erich B. RINGELSTEIN; Walter HUBER; Karl G. HAEUSLER; Stefanie BRUEHL; Klaus WILLMES; Frank DOMAHS; Frank REGENBRECHT; Klaus-Juergen SCHLENCK; Marion THOMAS; Ernst DELANGEN; Roman ROCKER; Franziska WIGBERS; Christina RUEHMKORF; Indra HEMPEN; Jonathan LIST; Hellmuth OBRIG; Arno VILLRINGER; Maria BLEY; Michael JOEBGES; Katja HALM; Joerg B. SCHULZ; Cornelius WERNER; Georg GOLDENBERG; Ralf GLINDEMANN; Gudrun KLINGENBERG; Eberhard KOENIG; Friedemann MUELLER; Berthold GROENE; Stefan KNECHT; Regina BAACKE; Janet KNAUSS; Stephanie MIETHE; Ulrich STELLER; Ralf SUDHOFF; Eva SCHILLIKOWSKI; Gustav PFEIFFER; Kathrin BILLO; Hannah HOFFMANN; Franz-Josef FERNEDING; Stephan RUNGE; Tina KECK; Volker MIDDELDORF; Stefan KRUEGER; Barbara WILDE; Karsten KRAKOW; Carla BERG-HOFF; Franziska REINHUBER; Ingeborg MASER; Werner E. HOFMANN; Christa SOUS-KULKE; Wilfried SCHUPP; Anke OERTEL; Detlef BAETZ; Farsin HAMZEI; Katja SCHULZ; Alfons MEYER; Angelika KARTMANN; O'Niel SOM; Solms-Bjoern SCHIPKE; Stephan BAMBORSCHKE.

Acknowledgements
The authors acknowledge the support of the Collaboration of Aphasia Trialists (CATs) which is funded by COST and The Tavistock Trust for Aphasia in fostering international aphasia research collaboration.

History
Manuscript accepted: June 23, 2025. - Manuscript received: February 19, 2025.

Supplementary data
For supplementary materials, please see the HTML version of this article at www.minervamedica.it