

Standardisierung der Auswertungskriterien des deutschsprachigen Amsterdam-Nijmegen Everyday Language Test (ANELT) zur Erhöhung der Beurteiler*innenübereinstimmung

I. Rubi-Fessen^{1,2}, A. Baumgärtner³, A. Flöel⁴, C. Breitenstein⁵

¹ Lehrstuhl für Pädagogik und Therapie bei Sprach- und Sprechstörungen, Department Heilpädagogik und Rehabilitation, Universität zu Köln

² Neurologische Rehabilitationsklinik RehaNova, Köln

³ Institut für Gesundheitswissenschaften und Studiengang Ergotherapie/Logopädie, Universität zu Lübeck

⁴ Klinik und Poliklinik für Neurologie, Universitätsmedizin Greifswald

⁵ Klinik für Neurologie mit Institut für Translationale Neurologie, Universität Münster

Zusammenfassung

Ein psychometrisch fundiertes Verfahren zur direkten Erfassung der verbalen Kommunikationsfähigkeit bei Aphasie ist der Amsterdam-Nijmegen Everyday Language Test (ANELT). Angesichts einer zunehmenden Verbreitung des ANELT in Forschung und Klinik ist eine Standardisierung der Auswertung wünschenswert. In der vorliegenden Untersuchung wurden, basierend auf den Antwortmustern von Personen ohne sprachliche Beeinträchtigungen, Auswertungskrite-

rien für die A-Skala des deutschsprachigen ANELT entwickelt und an Personen mit Aphasie überprüft. Das Bewertungsschema erwies sich bei den ersten Bewertungen von Patient*innenaufnahmen des ANELT als sehr hilfreich für die Gruppe der acht Beurteiler*innen. Eine formale Analyse der Beurteiler*innenübereinstimmung mittels Krippendorff's Alpha erwies sich als zufriedenstellend.

Schlüsselwörter: Aphasie, Diagnostik, verbale Kommunikation, psychometrische Gütekriterien

Einleitung

Personen mit Aphasie erfahren durch die Erkrankung häufig eine Einschränkung der Teilhabe am sozialen und beruflichen Leben. Kommunikative Alltagssituationen können oft nur mit Mühe bewältigt werden. Die Unfähigkeit, adäquat zu kommunizieren, wird von der Mehrheit der Betroffenen als »equal to or worse than death« bewertet [7]. In vielen Fällen kommt es zu einem starken sozialen Rückzug, der wiederum zu einer subjektiv wahrgenommenen verminderten Lebensqualität führt [8]. Neben der Befundung linguistischer Fähigkeiten gewinnt die Erfassung kommunikativ-pragmatischer Kompetenzen zunehmend an Bedeutung in der Aphasie-rehabilitation. Durch diese können die Auswirkungen der Erkrankung auf den Alltag von Personen mit Aphasie ökologisch valide abgebildet werden. Eine wiederholte Anwendung standardisierter kommunikativ-pragmatischer Testverfahren ermöglicht auch eine Erfassung der Veränderungen der Kommunikationsfähigkeit im Verlauf der Erkrankung.

Prinzipiell steht Personen mit Aphasie im Alltag neben der mündlichen Kommunikation auch die Kommunikation über Schrift, Gestik oder Zeichnungen zur Verfügung. Aus klinischer Erfahrung steht für viele Betroffene jedoch die Wiedererlangung der *verbalen*

Kommunikationsmöglichkeit im Fokus der Zielsetzung einer Aphasie-therapie.

Ein psychometrisch fundiertes Verfahren zur direkten Erfassung der verbalen Kommunikationsfähigkeit bei Aphasie ist der Amsterdam-Nijmegen Everyday Language Test (ANELT) [3]. Er untersucht die verbale Kommunikationsfähigkeit von Personen mit Aphasie im Quasi-Rollenspiel. Verfügbar sind zwei parallele Versionen [1] (Version I und II) mit jeweils 10 vorgegebenen Alltagssituationen, auf die verbal reagiert werden muss (z. B. telefonisch einen Arzttermin verlegen). Bei einem Teil der Situationen werden gebräuchliche Alltagsgegenstände einbezogen (z. B. Hemd, Handschuh und Korkenzieher). Die Äußerungen der Proband*innen werden anhand zweier 5-stufiger Skalen (A-Skala = inhaltliche Verständlichkeit, B-Skala = akustische Verständlichkeit) von externen Beurteiler*innen bewertet. Das ursprüngliche Vorgehen beim ANELT sieht ein »intuitives« Urteil der Bewerter*innen vor, das nicht von festgelegten Kriterien geleitet ist. Dabei wird für jede Situation die Äußerung der Proband*innen als »vollkommen verständlich« (Punktwert 5), »gut verständlich« (Punktwert 4), »mittelmäßig verständlich« (Punktwert 3), »wenig verständlich« (Punktwert 2) und »überhaupt nicht verständlich/keine Antwort« (Punktwert 1) klassifiziert. Bei der Beurteilung der A-Skala wird die formale Korrektheit (z. B.

phonologische oder syntaktische Angemessenheit) der Äußerung nicht in die Bewertung einbezogen. Es wird ausschließlich der kommunikationsrelevante Inhalt bewertet. Die für das niederländische Original berichtete sehr hohe Interrater Übereinstimmung (Krippendorff's $\alpha = 0,91$; 6 Beurteiler*innen, 14 Proband*innen) [3] wurde für eine deutsche Version des ANELT bislang nicht repliziert, jedoch pauschal als »hoch« [2] angegeben bzw. fiel selbst nach einer zeitaufwendigen Schulung der Beurteiler*innen unter Entwicklung eines Vorläufermodells des vorliegenden Auswertungsschema deutlich niedriger aus als für das niederländische Original (Krippendorff's $\alpha = 0,81$) [5].

Objektive Auswertungskriterien sollen den Anwender*innen helfen, Unsicherheiten bei der Bewertung zu begegnen. Angesichts einer zunehmenden Verbreitung des ANELT in Forschung und Klinik ist zudem eine Standardisierung der Auswertung wünschenswert. In der vorliegenden Untersuchung wurden, basierend auf den Antwortmustern von Personen ohne sprachliche Beeinträchtigungen, Auswertungskriterien für die A-Skala des

deutschsprachigen ANELT entwickelt und an Personen mit Aphasie überprüft.

Methode

Im Rahmen einer Bachelorarbeit [6] wurden die im ANELT produzierten Propositionen 40 älterer Personen ohne sprachliche Beeinträchtigungen (21 Frauen; Durchschnittsalter 75,5 Jahre; Streubreite 65–85 Jahre) erhoben und transkribiert. Die geäußerten Propositionen wurden für jede der 20 Situationen thematisch sortiert. Bei einem Teil der ANELT-Situationen zeigte sich eine hohe Übereinstimmung bezüglich der Äußerung spezifischer Propositionen (z.B. »Fernseher« beim Kauf eines entsprechenden Geräts). Bei anderen Situationen wurden Varianten von Propositionen mit vergleichbarer Bedeutung in einem Konsensprozess zwischen Erst- und Letztautorin zu übergeordneten Propositionen zusammengefasst (z.B. als »Schadensanzeige/Reklamation« beim Erhalt eines beschädigten Hemdes in der Wäscherei oder »Aktionsplan benannt« beim Optiker, wo mit Reparatur oder Neukauf einer beschädigten Brille verschiedene Handlungsalternativen möglich sind). Danach wurde pro Situation die Übereinstimmung der Teilnehmer*innen für jede der geäußerten (übergeordneten) Propositionen ausgezählt. Als »Kernproposition« wurden Äußerungen gewertet, die von mind. 90 Prozent der 40 Personen ohne sprachliche Beeinträchtigungen produziert wurden. Eine »Nebenproposition« musste von mindestens der Hälfte dieser Referenzgruppe geäußert werden. Zudem wurden weitere, inhaltlich passende Propositionen erfasst, die aber von weniger als der Hälfte der Personen ohne sprachliche Beeinträchtigungen genannt wurden. Neben dem Inhalt wurde auch die Anzahl der in der jeweiligen Situation hervorgebrachten Propositionen bestimmt.

Auf Grundlage dieser Kriterien (Kernpropositionen, Nebenpropositionen, inhaltlich passende Propositionen) wurde für jede der 20 Situationen festgelegt, welche Punktwerte der fünfstufigen ANELT-A Bewertungsskala (inhaltliche Verständlichkeit) für bestimmte Kriterienkonstellationen vergeben werden. Für 5 Punkte (vollkommen verständlich) müssen z.B. die Kernproposition(en) sowie zusätzlich mindestens eine der Nebenpropositionen der jeweiligen Situation geäußert werden. Ein Punkt (niedrigster Punktwert im ANELT) wird vergeben, wenn entweder keine Antwort erfolgt oder keinerlei inhaltlicher Bezug zur Situation erkennbar ist (Tab. 1 und Tab. 2). A) Das auf diese Weise erstellte Bewertungsschema wurde im Rahmen der DC-TRAIN-APHASIA Studie [10] mit acht Beurteiler*innen an mehreren Personen mit Aphasie evaluiert. Die acht Beurteiler*innen bewerteten nach einer Trainingsphase unabhängig voneinander die Tonaufnahmen dreier Personen mit Aphasie auf der ANELT A-Skala (insgesamt 20 Situationen der Versionen I und II). Die Höhe der Interrater-Übereinstimmung wurde jeweils

Tab. 1: Kern-, Neben- und fakultative Propositionen sowie Punkteschema am Beispiel »Wäscherei«

<p>Situation und Anweisung: »Sie sind jetzt in der Wäscherei und wollen das hier abholen und bekommen es so zurück« (Anmerkung: Hemd mit Brandloch wird gezeigt)</p>
<p>Kernproposition (KP, von mind. 90 % der Personen ohne sprachliche Beeinträchtigungen geäußert): 1) Schadensanzeige/Verwechslung, auch implizit (aber eindeutig) (100 %)</p>
<p>Nebenproposition (NP, von mind. 50 % der Personen ohne sprachliche Beeinträchtigungen geäußert): 2) Ersatzforderung/Verweigerung der Annahme (80 %)</p>
<p>Weitere von Personen ohne sprachliche Beeinträchtigungen geäußerte Propositionen 3) Benennen des Objekts (Oberhemd/Hemd, Bluse, »Ware«; 42,5 %) 4) Ursachenklärungswunsch (30 %) Gesamtzahl Propositionen: 90 % der Personen ohne sprachliche Beeinträchtigungen äußern mindestens zwei Propositionen</p>
<p>Punktvergabe</p> <ul style="list-style-type: none"> • 5 Punkte Kernproposition #1 + Nebenproposition (#2) • 4 Punkte: Kernproposition #1 • 3 Punkte: Nebenproposition (#2) ODER mind. 2 weitere inhaltlich relevante Propositionen • 2 Punkte: keine Kern-/Nebenproposition, aber minimaler Situationsbezug erkennbar • 1 Punkt: kein Bezug erkennbar

Tab. 2: Antwortbeispiele mit Punktvergabe am Beispiel »Wäscherei«

<p>Person 1 (ohne sprachliche Beeinträchtigungen, 5 Punkte): »Oh, da ist ja ein Brandfleck drin. Nee, die Ware nehme ich nicht zurück. Da möchte ich aber ein neues Hemd haben.«</p>
<p>Person 2 (mit Aphasie, 5 Punkte): »Ich möchte das nicht mehr – nicht kaufen. Das war eine Unverschämtheit. Dieser Fleck und dieser Brand. Das Das war ich nicht der Besitzer – sondern die Wäscherei.«</p>
<p>Person 3 (mit Aphasie, 4 Punkte): »De . ka . kas . äh .. se hat die kas die die . die nosch .. noscheko kasch .. de . de - mein Gott, was ist das dat Dings . der . der ha. der hat doch eine F-F-Fede . eine Fehler.«</p>
<p>Person 4 (mit Aphasie, 3 Punkte): »Hemd nein«</p>
<p>Person 5 (mit Aphasie, 2 Punkte): »Was hier?«</p>
<p>Person 6 (mit Aphasie, 1 Punkt): »Äh . das wäre . wäre . äh . doch nichts . das äh .. mhm .. das wäre doch nichts . das«</p>

mit Krippendorff's Alpha bestimmt. Die Berechnung von Krippendorff's Alpha als Maß für die Beurteiler*innenübereinstimmung erfordert streng genommen voneinander unabhängige Daten. Die wie im vorliegenden Fall bei einer einzelnen Person mit Aphasie erhobenen Bewertungen für 20 verschiedene ANELT Situationen sind vermutlich nicht unabhängig voneinander, da die Rater*innen ggf. dazu tendieren, für unterschiedliche Situationen einer Person mit Aphasie vergleichbare Werte zu vergeben. Da derzeit kein Verfahren zur Bestimmung der Beurteiler*innenübereinstimmung für abhängige Daten über mehrere Urteiler*innen hinweg verfügbar ist, wurde diese Limitation in Kauf genommen.

- B) Anhand des Therapieverlaufs zweier Personen mit Aphasie wurden zudem therapieinduzierte Veränderungen exemplarisch evaluiert. Dazu wurde der ANELT zu zwei Untersuchungszeitpunkten (U1 und U2) im Abstand von zwei Wochen erhoben. Dabei wurde Person A zu U1 mit der Version I und zu U2 mit der Version II untersucht, Person B zu U1 mit der Version II und zu U2 mit der Version I. Die Beurteilerin der anonymisierten Tonaufnahmen war bezüglich des Untersuchungszeitpunkts der Personen mit Aphasie jeweils verblindet. Beide Personen litten an einer postakuten amnestischen Aphasie und hatten zwischen den Untersuchungszeitpunkten zehn sprachtherapeutische Therapieeinheiten erhalten. Es wurde überprüft, ob die Veränderungswerte jeweils den in der ANELT A-Skala Einzelfalldiagnostik (bei Personen mit Aphasie in der postakuten Phase) anwendbaren kritischen Schwellenwert von >7 Punkten Unterschied [4] zwischen Vor- und Nachtestung überschreiten.

Ergebnisse

- A) Das Bewertungsschema erwies sich bei den ersten Bewertungen von Proband*innenaufnahmen des ANELT durchweg als sehr hilfreich für die acht Beurteiler*innen. Die Beurteiler*innenübereinstimmung (Krippendorff's Alpha) steigerte sich im Verlauf des Trainings und war insgesamt zufriedenstellend, insbesondere für die letzte ausgewertete Aufnahme (Person 3; Tab. 3) [9].

Tab. 3: Beurteiler*innenübereinstimmung für drei Probeaufnahmen

	Krippendorff's Alpha	95% KI untere Grenze	95% KI obere Grenze
Person 1	0,41	0,25	0,55
Person 2	0,52	0,47	0,57
Person 3	0,71	0,68	0,75

KI Konfidenzintervall

- B) Bei der Analyse der Verlaufsuntersuchungen zeigten sich für beide Personen mit Aphasie numerische

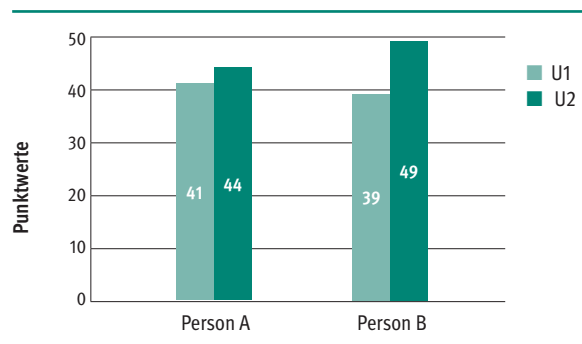


Abb. 1: : Veränderungen auf der ANELT A-Skala vor und nach Therapie

Verbesserungen. Person A erzielte einen Zuwachs von drei Punkten (von 41/50 Punkten in U1 auf 44/50 Punkte in U2), der aber unterhalb des kritischen Schwellenwertes in der ANELT-A-Skala lag. Person B überschritt mit einer Verbesserung von zehn Punkten (von 39/50 Punkten in U1 auf 49/50 Punkte in U2) den kritischen Schwellenwert für eine individuelle Verbesserung (>smallest detectable change; [10]) von 7 Punkten in der A-Skala (Abb. 1).

Diskussion

Die bisherigen Erfahrungen zeigen, dass die kriteriengeleitete Auswertung anhand des Bewertungsschemas leicht und anwenderfreundlich möglich ist und dass eine zufriedenstellende Beurteiler*innenübereinstimmung erzielt werden kann. Somit wird das standardisierte Bewertungsschema die Beurteilung des ANELT in Forschung und Klinik objektivieren und vereinheitlichen. Weitere Daten zur Beurteiler*innenübereinstimmung werden im Rahmen der DC-TRAIN-APHASIA Studie [11] erhoben werden.

Von den beiden Personen mit Aphasie, bei denen die Abbildbarkeit therapieinduzierter Veränderungen exemplarisch untersucht wurde, überschritt nur Person B den kritischen Schwellenwert von 7 Rohwertpunkten in der ANELT A-Skala. Neben der Kürze des Therapieintervalls von zwei Wochen könnte für die fehlende Signifikanz bei Person A auch die Untersuchungsreihenfolge der Versionen (II → I) eine Rolle gespielt haben, da die Vergleichbarkeit/Parallelität des Schwierigkeitsgrads der beiden Versionen der ANELT A-Skala (Version II eventuell schwieriger als Version I?) unter Kliniker*innen immer wieder diskutiert wird. Eine Analyse der Paralleltest-Reliabilität anhand der im Rahmen der FCET2EC-Studie [5] erhobenen Daten bei Personen mit chronischer Aphasie ergab jedoch keinen signifikanten Unterschied zwischen den Versionen I und II der A-Skala, wenn diese im Abstand von drei Wochen (ohne Studienintervention) abwechselnd durchgeführt wurden.

Eine Einschränkung der Studie ist der vergleichsweise hohe Altersdurchschnitt der Stichprobe der Personen ohne sprachliche Beeinträchtigungen. Dadurch, dass die

Neurol Rehabil 2022; 28(2): 57–60 | <https://doi.org/10.14624/NR2202001> | © Hippocampus Verlag 2022

Standardization of the evaluation criteria of the German-language Amsterdam-Nijmegen Everyday Language Test (ANELT) to increase inter-rater agreement

I. Rubi-Fessen, A. Baumgärtner, A. Flöel, C. Breitenstein

Abstract

A psychometrically sound procedure for the assessment of verbal communication skills in aphasia is the Amsterdam-Nijmegen Everyday Language Test (ANELT). In view of an increasing use of the ANELT in research and clinical practice, a standardization of the evaluation is needed. In the present study, based on the responses of persons without language impairments, evaluation criteria for the A scale of the German version of the ANELT were developed and tested on persons with aphasia. The scoring scheme proved to be very helpful for a group of eight raters during the first evaluations of patient recordings of the ANELT. A formal analysis of interrater agreement using Krippendorff's alpha demonstrated satisfactory results.

Keywords: aphasia, assessment, verbal communication, psychometric criteria

Situationen keinen generationsabhängigen Abruf festgelegter Propositionen/Wörter und Sätze erfordern, sondern eine sprachliche und formale Variabilität zulassen, ist jedoch ein alters- oder bildungsspezifischer Einfluss der Versprachlichung vermutlich vernachlässigbar.

Fazit und Ausblick

Durch die Anwendung des Bewertungsschemas in der DC-TRAIN-APHASIA Studie [10] wird die Veränderungssensitivität des Schemas an einer großen Stichprobe von Personen mit Aphasie evaluiert werden. Eine mögliche Normierung der ANELT A-Skala auf Grundlage des Bewertungsschemas muss in künftigen Studien erfolgen. Dazu sollte zudem an einer größeren Stichprobe überprüft werden, ob die kriteriengeleitete Auswertung der intuitiven Auswertung des Originals überlegen ist. Angesichts der häufigen Forderung klinischer Anwender*innen nach standardisierten Auswertungskriterien in der Aphasiediagnostik ist jedoch zu vermuten, dass der vorliegende Kriterienkatalog den Einsatz des ANELTs als routinemäßiges pragmatisch-kommunikatives Diagnostikverfahren in Klinik und Praxis erleichtert und ein sensitives Verfahren zur Erfassung von Therapieeffekten darstellt.

Literatur

1. Baumgärtner A, Honekamp A, Johannsen K, Kramer KN, Newesely G, Rubi-Fessen I et al. Standardisierung einer deutschsprachigen Version des Amsterdam-Nijmegen Everyday Language Test (ANELT) [Vortrag]. 14. Jahrestagung der Gesellschaft für Aphasieforschung und -behandlung (GAB), Freiburg, Deutschland, 7.11.2014
2. Blomert L, Buslach DC. Funktionelle Aphasiediagnostik mit dem Amsterdam-Nijmegen Everyday Language Test. *Forum Logopädie* 1994; 2: 3–6.
3. Blomert L, Kean ML, Koster C, Schokker J. Amsterdam-Nijmegen Everyday Language Test: construction, reliability and validity. *Aphasiology* 1994; 8(4): 381–407.
4. Blomert L, Koster C. Amsterdam-Nijmegen Test voor Alledagse Taalvaardigheden. Handleiding. Göttingen (DE): Hogrefe; 2008.

5. Breitenstein C, Grewe T, Flöel A, Springer L, Martus P, Huber W, et al. 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.
6. Buhler T, Scherer S. Wie antworten gesunde ältere Erwachsene auf die Items des Amsterdam-Nijmegen Everyday Language Test (ANELT)? Durchführung und Auswertung mit dem Ziel der Erhebung erster Normierungsdaten [Unveröffentlichte Bachelorarbeit]. Hamburg (DE): Hochschule Fresenius; 2014.
7. Everett EA, Everett W, Brier MR, White P. Appraisal of health states worse than death in patients with acute stroke. *Neurol Clin Pract* 2021; 11(1): 43–8.
8. Hilari K. The impact of stroke: are people with aphasia different to those without? *Disabil Rehabil* 2011; 33(3): 211–8.
9. Hughes J. krippendorffsalpha: An R Package for Measuring Agreement Using Krippendorff's Alpha Coefficient. *R Journal* 2021; 13(1): 413–25. weblink: <https://journal.r-project.org/archive/2021/RJ-2021-046/RJ-2021-046.pdf>
10. Mokkink LB, Boers M, van der Vleuten CPM, 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.
11. Stahl B, Darkow R, von Podewils V, Meinzer M, Grittner U, Reinhold T, et al. Transcranial direct current stimulation to enhance training effectiveness in chronic post-stroke aphasia: A randomized controlled trial protocol. *Front Neurol* 2019; 10: 1089.

Danksagung: Wir bedanken uns bei A. Araújo da Cunha für die Unterstützung bei der Systematisierung der Propositionen und bei K. Jonas für die Bewertung der Tonaufnahmen zur Messung der therapieinduzierten Veränderungen. Dank auch an das DC-TRAIN-APHASIA »endpoint committee« mit M. Abel, A. Honekamp, K. Jonas, C. Kähler, G. Newesely, I. Rubi-Fessen, B. Wachtlin und Ch. Wendt für die fachlichen Diskussionen.

Interessenvermerk

Bei keiner der Autorinnen besteht ein Interessenskonflikt.

Korrespondenzadresse:

Dr. rer. medic. Ilona Rubi-Fessen
Universität zu Köln
Humanwissenschaftliche Fakultät
Pädagogik und Therapie bei Sprach- und Sprechstörungen
Klosterstraße 79b
50931 Köln
ilona.rubi-fessen@uni-koeln.de