Taylor & Francis
Taylor & Francis Group

ORIGINAL ARTICLE

🔓 OPEN ACCESS Check for updates

German version of the Chelsea Critical Care Physical Assessment Tool (CPAx-GE): translation, cross-cultural adaptation, validity, and reliability

Sabrina Eggmann[a,b] 🆔, Martin L. Verra[a], Valentine Stefanicki[c], Angela Kindler[a], Daphne Seyler[a], Roger Hilfiker[d], Joerg C. Schefold[e], Caroline H. G. Bastiaenen[b*] and Bjoern Zante[e*]

[a]Department of Physiotherapy, Inselspital, Bern University Hospital, Bern, Switzerland; [b]Department of Epidemiology, Research Line Functioning, Participation and Rehabilitation CAPHRI, Maastricht University, Maastricht, The Netherlands; [c]Physiotherapy, Swiss Paraplegic Centre, Nottwil, Switzerland; [d]School of Health Sciences, HES-SO Valais-Wallis, Leukerbad, Switzerland; [e]Department of Intensive Care Medicine, Inselspital, Bern University Hospital, University of Bern, Bern, Switzerland

ABSTRACT

Purpose: To translate and cross-culturally adapt the Chelsea Critical Care Physical Assessment tool from English to German (CPAx-GE) and to examine its validity and reliability.

Materials and methods: Following a forward-backward translation including an expert round table discussion, the measurement properties of the CPAx-GE were explored in critically ill, mechanically ventilated adults. We investigated construct, cross-sectional, and cross-cultural validity of the CPAx-GE with other measurement instruments at pre-specified timepoints, analysed relative reliability with intraclass correlation coefficients (ICCs) and determined absolute agreement with the Bland–Altman plots.

Results: Consensus for the translated CPAx-GE was reached. Validity was excellent with >80% of the pre-specified hypotheses accepted at baseline, critical care, and hospital discharge. Interrater reliability was high (ICCs > 0.8) across all visits. Limit of agreement ranged from −2 to 2 points. Error of measurement was small, floor, and ceiling effects limited.

Conclusions: The CPAx-GE demonstrated excellent construct, cross-sectional, and cross-cultural validity as well as high interrater reliability in critically ill adults with prolonged mechanical ventilation at baseline, critical care, and hospital discharge. Consequently, the CPAx-GE can be assumed equal to the original and recommended in the German-speaking area to assess physical function and activity of critically ill adults across the critical care and hospital stay.

Trial registration: German Clinical Trials Register (DRKS) identification number: DRKS00012983 (https://www.drks.de/drks_web/navigate.do?navigationId=trial.HTML&TRIAL_ID=DRKS00012983), registered on 20 September 2017, first patient enrolled on 21 November 2017.

➤ **IMPLICATIONS FOR REHABILITATION**
- Early rehabilitation of critically ill patients is recommended to prevent and treat the subsequent functional disability, but a suitable measurement instrument for the German-speaking area is lacking.
- The translated, cross-culturally adapted German CPAx demonstrated excellent validity and reliability in assessing physical function and activity in critically ill adults.
- Cross-sectional validity of the CPAx has been newly established and allows the use of this tool at clinically relevant time-points in the course of a critical illness.
- The CPAx-GE can therefore be used in clinical practice by German-speaking therapists to assess physical function and activity during early rehabilitation in the ICU and hospital.

Introduction

Critical illness with prolonged mechanical ventilation is associated with loss of muscle mass accompanied by profound, generalised muscle weakness that has been termed intensive care unit acquired weakness (ICU-AW) [1–4]. Recovery is often prolonged and incomplete [5,6], but seems to be facilitated by early rehabilitation starting in the ICU [7,8]. However, to properly assess, treat, and evaluate the physical function and activity of critically ill patients a practical, valid, and reliable assessment tool is needed [9].

The Chelsea Critical Care Physical Assessment tool (CPAx) has been specifically developed in the UK to measure physical function and activities in the ICU [10,11]. The CPAx is a performance-

based measurement instrument based on observation. The 10 items evaluate respiratory function, functional mobility (position changes, sitting, standing, stepping), and grip strength with an ordinal scale from 0 (unable/dependent) to 5 (independent) resulting in a maximum score of 50 (independent) and a minimum score of 0 (completely dependent). The measured constructs of the CPAx are allocated to the domains of body functions and activities – particularly muscle functions, movement functions, functions of the respiratory system and mobility – according to the International Classification of Functioning, Disability, and Health (ICF) [12]. A recent systematic review confirmed the CPAx as a robust measurement for critically ill patients with excellent measurement properties, although cross-sectional validity, i.e., the validity at different time-points in the course of a disease, has not yet been established [13]. The CPAx has not yet been officially translated and cross-culturally adapted into German. To make the measurement instrument available to the German-speaking area, it is therefore necessary to establish a conceptually equivalent version to the original CPAx that considers the German-speaking cultural context [14].

The current study aimed to translate and cross-culturally adapt the CPAx from English into German (CPAx-GE) and then to investigate construct, cross-sectional and cross-cultural validity, interrater reliability, internal consistency and measurement error of the translated CPAx-GE at pre-specified timepoints across the ICU and hospital stay in German-speaking Switzerland. We hypothesised that the CPAx-GE would primarily have a moderate to high correlation with the Medical Research Council sum score (MRC-SS) – the standard for diagnosing an ICU-AW – across the pre-specified visits and thus be comparable to the original version. We further expected the CPAx-GE to have high interrater reliability (intraclass correlation coefficients (ICCs) >0.8) and good internal consistency (Cronbach's $\alpha > 0.7$) over all visits.

Materials and methods

Translation and cross-cultural adaptation

After receiving permission to translate the CPAx into German from the original developer (Dr. Evelyn J. Corner), the CPAx was translated using international recommended standards of a stepwise, forward-backward approach including cross-cultural adaptation with a multidisciplinary expert committee [14,15]. The rigorous process is illustrated in Figure 1. The expert committee included the four bilingual translators, a project manager, a medical doctor, a teacher, and an epidemiologist. The relevant cultural origins of the committee were Switzerland, Germany, UK, and Ireland. Following consensus of the expert committee meeting, the preliminary German CPAx version was tested pragmatically [14–16] by six independent physiotherapists for practicability, feasibility, and content validity during routine physiotherapy care (Figure 1). Their feedback aimed to inform the expert committee about the practicability and comprehensibility of the tool to create the final CPAx-GE version.

Design and setting

The final and approved CPAx-GE version was tested in a prospective, single-centre, longitudinal, clinimetric study that was conducted in a large, interdisciplinary ICU of an academic hospital (Department of Intensive Care Medicine, Inselspital, Bern University Hospital, Switzerland). The longitudinal study design aimed to assess the measurement properties of the CPAx within the entire trajectory of the critically ill patient at pre-specified

timepoints from ICU baseline to hospital discharge. The study was approved by the local Ethics Committee of Canton Berne on 14 August 2017 (ID 2017-01396) and was prospectively registered in the German Clinical Trials Register (DRKS00012983) on 20 September 2017. The study was therefore subject to the Declaration of Helsinki, Swiss federal law and Good Clinical Practice (GCP) guidelines. Trial reporting follows the STROBE statement [17] and uses COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) methodology [9].

Participants and procedures

The CPAx was designed for ICU patients at risk of ICU-AW and its subsequent morbidity [10]. Accordingly, the inclusion criteria were an ICU stay with ≥72 h of mechanical ventilation, age >18 years and sufficient language skills in oral and written German. Exclusion criteria were planned discharge within the next 24 h, a strictly neurological ICU admission diagnosis, ongoing palliative treatment, pre-existing mental disability, participants living in a care facility, wheelchair users, previous external ICU stay for more than three days, ICU re-admission, or participation in ≥2 studies.

Assessors were required to have qualified as a physiotherapist and to conduct the official CPAx training (<2 h, in English) at http://cpax.ocbmedia.com. They further received a handout with study procedures that was explained by a co-author (SE, AK, DS). All physiotherapists spoke German fluently and originated from either Switzerland or Germany. Experience ranged from recently qualified to >10 years' experience in the ICU. The CPAx was not used at this institution prior to the study.

Daily screening of eligible ICU patients was conducted by a research nurse in consultation with a senior physiotherapist. Prior to study inclusion, eligible patients' proxies were approached by a GCP-trained physiotherapist to obtain written informed consent. As soon as participants regained the capacity to decide for themselves, written informed consent from the patient was sought. If this capacity was not regained within 90 days after ICU discharge, the ethics committee waived further patient consent. Participants were only included between 72 and 144 h of mechanical ventilation to avoid delayed baseline assessments.

Validity

The COSMIN panel defines validity as "the degree to which an instrument measures the construct(s) it purports to measure" [9]. Accordingly, we formulated *a priori* hypotheses about the relationship of the CPAx-GE with other instruments measuring convergent (related) or divergent (not related) constructs as described by the ICF framework (construct validity). These 22 hypotheses were based – when available – on the previous performance of the original CPAx (cross-cultural validity) [10]. Additionally, we used our understanding of the underlying constructs (see newly created conceptual model in Supplementary Material 1) at the three presumed most relevant visits (V1: baseline, V3: ICU discharge, V4: hospital discharge) in the pathway of the critically ill (cross-sectional validity). The hypotheses and their underlying constructs are outlined in Supplementary Material 1. Our assumption was that the relationship between the CPAx and other measurement instruments would differ between the three chosen timepoints in the course of a critical illness due to temporal changes in participants' measured characteristics. Accordingly, cross-sectional validity refers to single scores at a selected point in time [9] and was considered an important measurement property to investigate in the target population. Validity
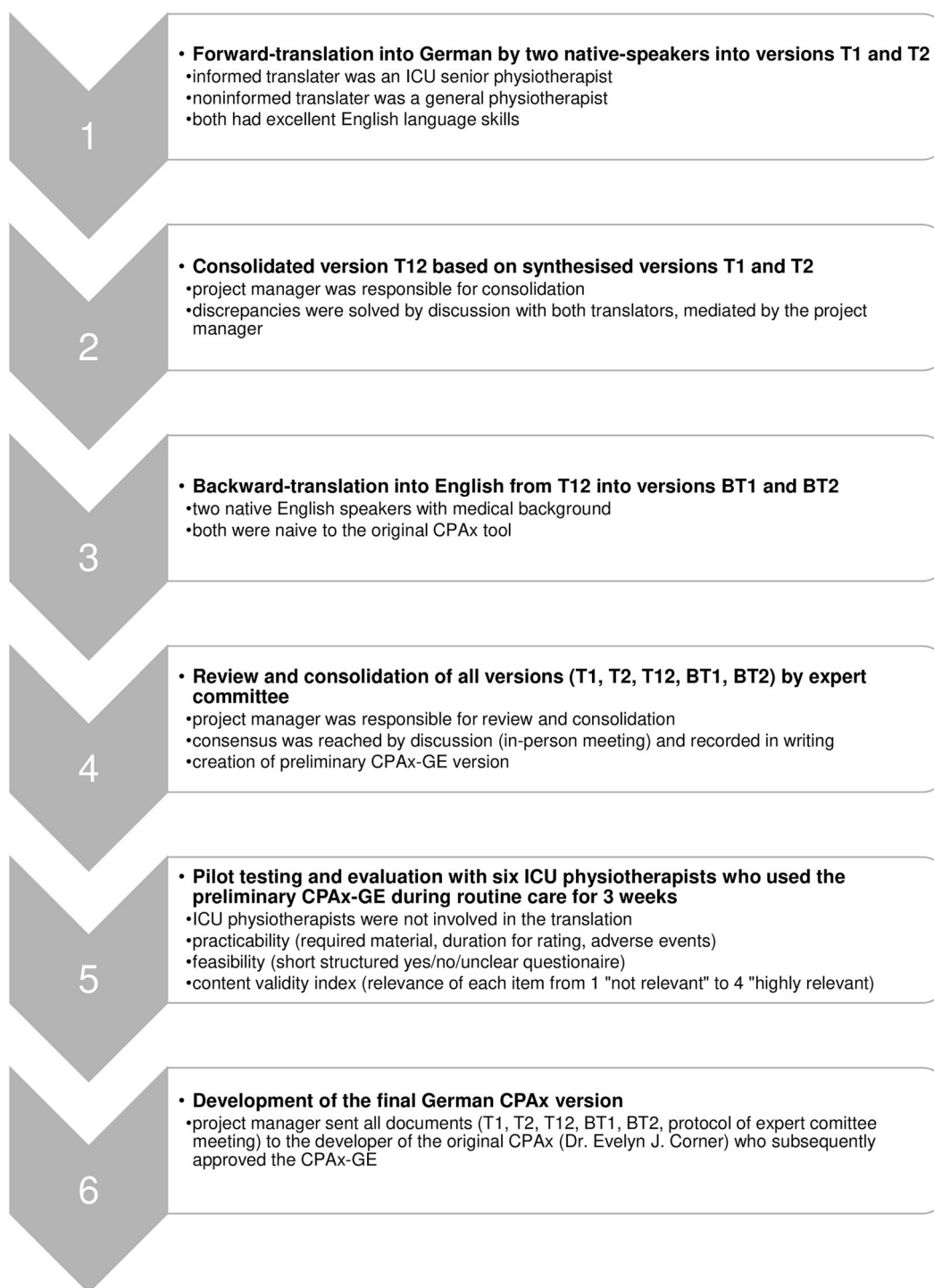
1
- **Forward-translation into German by two native-speakers into versions T1 and T2**
  - informed translater was an ICU senior physiotherapist
  - noninformed translater was a general physiotherapist
  - both had excellent English language skills

2
- **Consolidated version T12 based on synthesised versions T1 and T2**
  - project manager was responsible for consolidation
  - discrepancies were solved by discussion with both translators, mediated by the project manager

3
- **Backward-translation into English from T12 into versions BT1 and BT2**
  - two native English speakers with medical background
  - both were naive to the original CPAx tool

4
- **Review and consolidation of all versions (T1, T2, T12, BT1, BT2) by expert committee**
  - project manager was responsible for review and consolidation
  - consensus was reached by discussion (in-person meeting) and recorded in writing
  - creation of preliminary CPAx-GE version

5
- **Pilot testing and evaluation with six ICU physiotherapists who used the preliminary CPAx-GE during routine care for 3 weeks**
  - ICU physiotherapists were not involved in the translation
  - practicability (required material, duration for rating, adverse events)
  - feasibility (short structured yes/no/unclear questionaire)
  - content validity index (relevance of each item from 1 "not relevant" to 4 "highly relevant")

6
- **Development of the final German CPAx version**
  - project manager sent all documents (T1, T2, T12, BT1, BT2, protocol of expert comittee meeting) to the developer of the original CPAx (Dr. Evelyn J. Corner) who subsequently approved the CPAx-GE

Figure 1. Step-wise translation and cross-cultural adaptation based on [14,15].

was considered excellent if ≥75% of the hypotheses were accepted at each visit [18]. The measured instruments and their constructs as well as the time of their assessment are described in Table 1 [19–25].

Reliability

Reliability means "the degree to which a measurement is free from measurement error" [9]. Interrater reliability thereby refers to different persons rating the instrument on the same occasion [26]. Given that ICU patients' status may change rapidly, the CPAx-GE was rated after a standard physiotherapy session by trained physiotherapists ($n = 11$). This session was performed and rated by the treating physiotherapist while one to two raters scored the CPAx-GE by observation. Interrater(s) were chosen by availability and were blinded to each other's scores. Initially, interrater reliability assessments were randomly allocated to a subgroup of patients ($n = 20$) at one of the four visits (V1–4). However, to increase the sample size, the local ethics committee approved an amendment after 16 patients so that each

Table 1. Description of measured instruments, their constructs, and schedule.

Measurement instrument	Description, validity, and reliability	Construct (ICF component)	Visits[a]
Medical Research Council sum score (MRC-SS)[b]	Measures strength on a scale from 0 (no contraction) to 5 (normal strength) in 6 bilateral muscle groups, maximal summed score of 60 Reference standard to measure ICUAW (MRC-SS <48) [1], not validated Performance-based, volitional test (participants who were unable to follow at least 3 out of 5 commands, were assessed based on observation during physiotherapy) Established reliability in responsive, critically ill patients (ICC = 0.95) [19]	Muscle functions (b730–b749)	V1, V2, V3, V4
Functional Independence Measure (FIM)[b]	18-item scale that evaluates assistance in activities of daily living such as self-care, sphincter control, locomotion, or communication on a scale from 1 (completely dependent) to 7 (completely independent) Multidimensional, performance-based, observational No established reliability and validity in the critically ill patient	Diverse body functions, activities, and participation	V1, V3, V4
ICU Mobility Scale (IMS)[b]	Describes the highest level of mobility in critically ill adults from 0 (lying in bed with no active movement) to 10 (walking independently without aids) Performance-based, observational Good interrater reliability (weighted Kappa 0.83) [20] and established construct and predictive validity [21]	Mobility (d4)	V1, V2, V3, V4
Peak Cough Flow (PCF)[b]	Assessed with a peak flow meter (Vitalograph Model 4300, Vitalograph Ltd., Ennis, Ireland) during volitional coughing in extubated or tracheotomised participants who were able to follow at least 3 out of 5 commands Performance-based, volitional No established reliability and validity in the critically ill patient	Additional respiratory functions (b450)	V1, V3, V4
Bogenhausener Dysphagia sum-score (BODS)[b]	Rates dysphagia severity from 2 (no dysphagia) to 16 (severe dysphagia) by evaluating saliva swallowing and oral intake Performance-based, observational No established reliability and validity in the critically ill patient	Swallowing (b5105)	V1, V3, V4
Sequential Organ Failure Assessment Score (SOFA)[c]	Describes the degree of organ dysfunction from 0 (normal) to 4 (most abnormal) for six vital organs Observational (biological and physiological variables) Content validity established by a consensus meeting [22], excellent interrater reliability (ICC = 0.89) [23]	Body structures	V1, V2, V3
Glasgow Coma Scale (GCS)[c]	Assessment of consciousness since 1974 Performance-based, observational Reliability has been summarised as good and its worldwide use validated its relevance in neurologically impaired patients [24]	Mental functions (b1), movement functions (b750–b789)	V1, V2, V3
Richmond Agitation Sedation Scale (RASS)[c]	Used to titrate sedation ranging from unarousable (−5) to calm and alert (0) to combative (+4) Performance-based, observational confirmed construct validity and high interrater reliability (ICC > 0.9) [25]	Mental functions (b1)	V1, V2, V3

[a]V1: baseline, assessed between day 4 and 7; V2: ICU stay, assessed from day 5 every Tuesday and Friday until ICU discharge or a maximum of 10 assessments; V3: ICU discharge, assessed at ICU discharge; V4: hospital discharge, assessed at hospital discharge.
[b]Performed or collected routinely by trained physiotherapists.
[c]Extracted from routinely collected data via Patient Data Management System; the SOFA is routinely assessed daily at approximately 6 am, GCS and RASS were extracted at the time before physiotherapy.

participant would be rated at ICU discharge (V3) by at least two raters. This timepoint was considered to be the most important for future randomised controlled trials in critical care rehabilitation because it might reveal immediate effects on physical function and activity. We expected high ICCs of >0.8 based on the original CPAx [10].

Statistical analysis

Demographic characteristics and distribution of scores were analysed with descriptive statistics. We considered floor and ceiling effects to be acceptable if ≤15% [18]. The CPAx is an ordinal scale [10], therefore, data are presented as a median with an interquartile range (IQR) or as numbers with percentages and analysed with nonparametric statistics. Analyses were performed with SPSS (IBM SPSS Statistics Version Premium GradPack 24, Armonk, NY) and R (Version 3.6.1, Vienna, Austria).

Validity-hypotheses were tested with Spearman's rank correlation coefficient (total score-based, except for item-based hypotheses). Sample size calculation with a power of $\beta = 0.8$, $\alpha = 0.05$, and a correlation of $r > 0.5$ between the MRC-SS and the CPAx

warranted a sample of 29 patients [27]. To compensate for drop-outs, the targeted sample size was increased to $n = 60$.

ICCs with 95% confidence intervals (CIs) were analysed by absolute agreement with a one-way random, single measurement model [ICC(1,1)] because not every subject was rated by each rater as suggested by [28] for each visit (V1–4, total score-based) and across all visits (total score and item-based) using a linear mixed model [29]. The Bland–Altman plots [30] comparing the treating versus the observing physiotherapist were constructed for each visit. Limit of agreement across all visits was adjusted for repeated measurements with the modified "true value varies" method [31]. Internal consistency was calculated with Cronbach's alpha for single visits to compare results with the original CPAx version ($\alpha > 0.7$). Finally, standard error of measurement (SEM) by agreement and smallest detectable change (SDC = ±1.96× $\sqrt{2} \times$ SEM) were investigated for single and total visits (V1–4) [32]. Based on the original study [10], we primarily considered a random subgroup ($n = 20$) to assess interrater reliability over all timepoints as adequate. However, the minimal recommended sample size for reliability studies is $n = 50$ [18], which led to an amendment after 16 patients as discussed above.
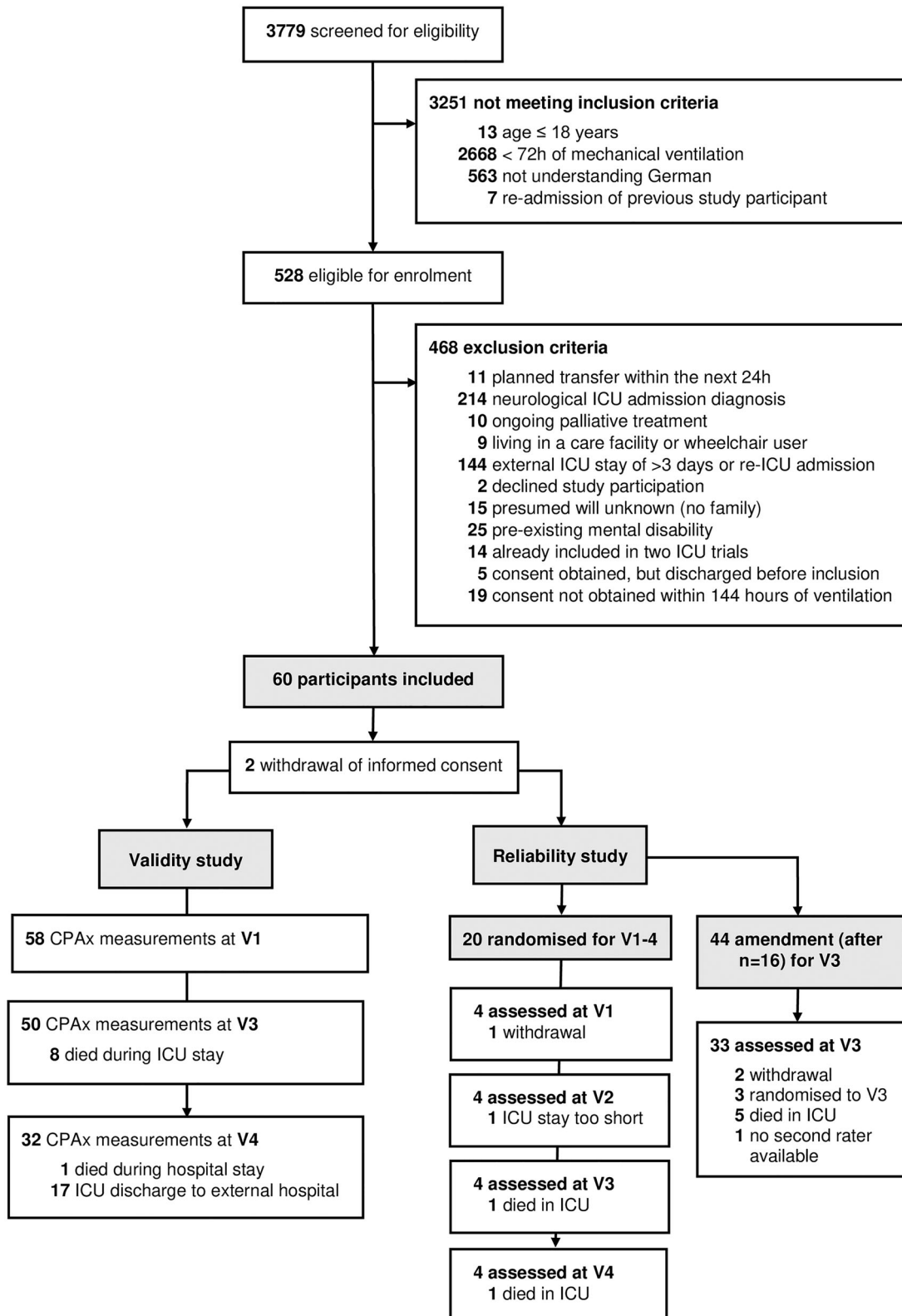
3779 screened for eligibility

3251 not meeting inclusion criteria

- **13** age ≤ 18 years
- **2668** < 72h of mechanical ventilation
- **563** not understanding German
- **7** re-admission of previous study participant

528 eligible for enrolment

468 exclusion criteria

- **11** planned transfer within the next 24h
- **214** neurological ICU admission diagnosis
- **10** ongoing palliative treatment
- **9** living in a care facility or wheelchair user
- **144** external ICU stay of >3 days or re-ICU admission
- **2** declined study participation
- **15** presumed will unknown (no family)
- **25** pre-existing mental disability
- **14** already included in two ICU trials
- **5** consent obtained, but discharged before inclusion
- **19** consent not obtained within 144 hours of ventilation

60 participants included

2 withdrawal of informed consent

Validity study

58 CPAx measurements at **V1**

50 CPAx measurements at **V3**
- **8** died during ICU stay

32 CPAx measurements at **V4**
- **1** died during hospital stay
- **17** ICU discharge to external hospital

Reliability study

20 randomised for V1-4

4 assessed at V1
- **1** withdrawal

4 assessed at V2
- **1** ICU stay too short

4 assessed at V3
- **1** died in ICU

4 assessed at V4
- **1** died in ICU

44 amendment (after n=16) for V3

33 assessed at V3
- **2** withdrawal
- **3** randomised to V3
- **5** died in ICU
- **1** no second rater available

Figure 2. Study flow. V1: visit at baseline; V2: visit at ICU stay; V3: visit at ICU discharge; V4: visit at hospital discharge.

Table 2. Demographics and characteristics of study participants ($n = 58$).

Variables	Participants
Age (years)	68 [56, 73]
Gender (male)	41 (70.7%)
BMI (kg/m^2)	26 [22.8, 29.1]
Weight (kg)	75.5 [70, 94]
APACHE II score (0–71)[a]	32 [28, 36]
SOFA score (0–24)[a]	12 [10, 14]
ICU diagnosis on ICU admission	
Cardiac surgery	15 (25.9%)
Other surgery	7 (12.1%)
Respiratory insufficiency	15 (25.9%)
Hemodynamic insufficiency	13 (22.4%)
Trauma	5 (8.6%)
Neurology/neurosurgery	–
Gastroenterology	1 (1.7%)
Other	2 (3.4%)
ICU days until study inclusion	3.84 [3.41, 4.70]
Length of stay in ICU (days)[b]	7.97 [6.69, 12.85]
Duration of mechanical ventilation (days)[b]	6.81 [5.30, 11.51]
Tracheostomy[b]	15 (25.9%)
ICU readmission[b]	7 (12%)
Length of stay in hospital (days)[b]	21.08 [11.72, 27.18]
ICU mortality[b]	8 (13.8%)
ICU and hospital mortality[b]	9 (15.5%)
ICU discharge destination	
Hospital wards	34 (58.6%)
External ICU	13 (22.4%)
External hospital	3 (5.2%)
Died	8 (13.8%)
Hospital discharge destination	
Rehabilitation	24 (41.4%)
External hospital	20 (34.5%)
Home	5 (8.5%)
Died	9 (15.5%)

BMI: body mass index; APACHE: Acute Physiology and Chronic Health Evaluation; SOFA: Sequential Organ Failure Assessment; ICU: intensive care unit. Data are presented as median [25%, 75%] or frequencies n (%).
[a]ICU admission.
[b]At original hospital (excluding external hospitals).

Results

Translation and cross-cultural adaptation

The CPAx-GE was created following the step-wise protocol in Figure 1. The expert committee reviewed all items of the five translated versions (T1, T2, T12, BT1, BT2) and discussed several minor cultural and linguistic differences (Supplementary Material 2). All members agreed on the preliminary German CPAx version to be tested in clinical practice. In this pilot trial, the preliminary CPAx-GE was found to be feasible and practicable with an excellent content validity (Supplementary Material 2). One minor uncertainty about the scientific unit of oxygen was revealed and subsequently changed by the expert committee to litres/minute. Afterwards the final German CPAx (Supplementary Material 3) was created and the cultural adaptations approved by the original developer.

Validity and reliability

From 21 November 2017 to 25 May 2019, 528 eligible patients were screened for study inclusion of whom 60 had no exclusion criteria and were included in the study (Figure 2). Two informed consents were withdrawn after study inclusion, we therefore report the results of 58 critically ill patients. Their demographics and characteristics are presented in Table 2. Floor (10%) and ceiling effects (6%) were highly acceptable for the CPAx across all three visits, especially when compared to other measures of mobility and activity (Supplementary Material 4).

Construct and cross-sectional validity of the CPAx-GE were excellent with 86% (19 out of 22) of overall hypotheses and $\geq$80% of hypotheses at each time-point (V1, V3, V4) accepted (Table 3). The acceptance rate of the cross-cultural hypotheses – based on the original CPAx – was 83% (10 out of 12) (Table 3).

Interrater reliability by observation was excellent with ICCs >0.8 including 95% CI for single and combined, adjusted visits (V1–4) (Table 4). The constructed Bland–Altman's plots confirmed the high agreement of the CPAx-GE with a mean difference of 0.13 ± 0.15 (95% limit of agreement: -2.04 to 1.79) across all visits (Figure 3) or lower for single visits (Supplementary Material 4). On the CPAx-GE scale, the limit of agreement therefore ranges from -2 to 2 (four points). Cronbach's alpha was comparable to the original CPAx with $\alpha > 0.7$ for V1–V3, but not V4 ($\alpha = 0.61$). Across all visits, the SEM and SDC for the CPAx-GE were 0.68 and 1.89, respectively (Table 4).

Discussion

A final German CPAx version was developed and approved by the original developer as a conceptually equivalent version to the original CPAx. The CPAx-GE demonstrated excellent construct, cross-sectional, and cross-cultural validity across the ICU and hospital stay in critically ill adults with prolonged mechanical ventilation (>72 h). Similarly, floor and ceiling effects were very acceptable (<15%) across the entire trajectory. Interrater reliability was high with ICCs >0.8 at all four visits. The 95% limit of agreement as identified by Bland–Altman's plotting was four points, which is lower than the established minimal clinically important difference of six points [33]. Internal consistency ($\alpha > 0.7$) over the ICU stay was comparable to the original CPAx indicating consistent cross-cultural measurement. Finally, measurement errors expressed by SEM (one point) and SDC (two points) were low enhancing interpretability of the CPAx-GE.

This study meticulously followed the suggested international standards for the translation and cross-cultural adaptation of a measurement instrument [14,15]. Discrepancies were solved by discussion within a multidisciplinary expert committee and verified with the original author to ensure a conceptually equivalent version. This approach has proved to be successful for the currently available Swedish and Danish CPAx translations [34,35]. Nevertheless, there are limitations to the German translation. While we were careful to include people with a German background in the translation process, we cannot ensure cross-cultural validity to all German-speaking regions because all members of the expert committee were currently living in Switzerland. Additionally, the preliminary German CPAx version was tested in a single-centre in Switzerland and thus may not be generalisable to other regions.

The subsequent clinimetric study rigorously employed the methodological principles of the COSMIN initiative [9]. Similar to the validation of the original CPAx [10], we investigated validity using hypothesis-testing to compare the relationships of other measurement instruments with partly overlapping theoretical constructs. This included the same measurement instruments as in the original study (e.g., MRC-SS, SOFA, GCS, and PCF), along with specifically included new measurements of mobility (e.g., IMS) to expand our knowledge on the underlying theoretical constructs. In contrast to the original study [10], we did not assess these measurements at a random timepoint during the ICU stay, but we specifically formulated theory-driven hypotheses across

Table 3. Hypothesis-testing for construct, cross-sectional, and cross-cultural validity of the CPAx-GE.

No.	Prior hypothesis	n	Spearman's rank correlation coefficient (r)	Confirmed
ICU baseline (V1), n = 58				
1	CPAx with MRC-SS (r = 0.5–0.8)[a]	58	0.639	Yes
2	CPAx with FIM (r = 0.4–0.7)	58	0.554	Yes
3	CPAx with IMS (r = 0.6–0.9)	58	0.711	Yes
4	CPAx with BODS (r = 0 to −0.3)	58	−0.294	Yes
5	CPAx with SOFA (r = −0.4 to −0.7)[a]	58	−0.341	No
6	CPAx with GCS (r = 0.5–0.8)[a]	58	0.603	Yes
7	CPAx with RASS (r = 0.3–0.6)[a]	58	0.485	Yes
Total of accepted hypotheses for V1: 6 out of 7 (86%)				
ICU discharge (V3), n = 50				
8	CPAx with MRC-SS (r = 0.6–0.9)[a]	50	0.809	Yes
9	CPAx item 10 with MRC-SS (r > 0.7)	50	0.775	Yes
10	CPAx with FIM (r > 0.7)[a]	50	0.811	Yes
11	CPAx with IMS (r > 0.7)	50	0.883	Yes
12	CPAx item 2 with PCF (r = 0.4–0.7)[a]	35	0.524	Yes
13	CPAx with BODS (r = −0.3 to −0.6)	50	−0.540	Yes
14	CPAx with SOFA (r = −0.4 to −0.7)[a]	50	−0.482	Yes
15	CPAx with GCS (r = 0.5–0.8)[a]	50	0.781	Yes
16	CPAx with RASS (r = 0.2–0.4)	50	0.082	No
Total of accepted hypotheses for V3: 8 out of 9 (89%)				
Hospital discharge (V4), n = 32				
17	CPAx with MRC-SS (r = 0.5–0.8)[a]	31	0.674	Yes
18	CPAx item 10 with MRC-SS (r > 0.7)	31	0.391	No
19	CPAx with FIM (r > 0.7)[a]	32	0.820	Yes
20	CPAx with IMS (r > 0.7)	32	0.856	Yes
21	CPAx item 2 with PCF (r = 0.4–0.7)[a]	27	0.432	Yes
22	CPAx with BODS (r = 0 to −0.3)	32	0.027	Yes
Total of accepted hypotheses for V4: 5 out of 6 (83%)				
Total of accepted hypotheses for complete pathway: 19 out of 22 (86%)				

MRC-SS: Medical Research Council sum score; FIM: Functional Independence Measure; IMS: ICU Mobility Scale; BODS: Bogenhausener Dysphagia sum-score (BODS); SOFA: Sequential Organ Failure Assessment Score; GCS: Glasgow Coma Scale; RASS: Richmond Agitation Sedation Scale; PCF: Peak Cough Flow.
[a]Cross-cultural validity: hypothesis based on original CPAx [10].

Table 4. Intraclass correlation coefficients (ICCs) for interrater reliability of the CPAx-GE.

Visit	Item	ICC (95% CI)	SEM	SDC	Cronbach's α
V1 (ICU baseline)	Total CPAx score	0.984 (0.85–0.999)	0.500	1.386	0.85
V2 (ICU stay)	Total CPAx score	0.998 (0.991–1)	0.483	1.339	0.94
V3 (ICU discharge)	Total CPAx score	0.995 (0.992–0.997)	0.688	1.907	0.93
V4 (Hospital discharge)	Total CPAx score	0.979 (0.809–0.999)	0.500	1.386	0.61
Combined (V1–4)[a]	Total CPAx score	0.996 (0.995–0.998)	0.680	1.885	
Combined (V1–4)[a]	Respiratory function	0.919 (0.914–0.948)	0.303	0.840	
Combined (V1–4)[a]	Cough	0.993 (0.993–0.993)	0.102	0.283	
Combined (V1–4)[a]	Moving within the bed	0.902 (0.866–0.951)	0.297	0.823	
Combined (V1–4)[a]	Supine to sitting on the edge of the bed	0.970 (0.971–0.974)	0.237	0.657	
Combined (V1–4)[a]	Dynamic sitting	0.976 (0.974–0.978)	0.236	0.654	
Combined (V1–4)[a]	Standing balance	0.974 (0.963–0.983)	0.181	0.502	
Combined (V1–4)[a]	Sit to stand	0.967 (0.957–0.973)	0.291	0.807	
Combined (V1–4)[a]	Transfers	0.967 (0.959–0.968)	0.333	0.923	
Combined (V1–4)[a]	Stepping	0.968 (0.953–0.962)	0.291	0.807	
Combined (V1–4)[a]	Grip strength	0.995 (0.995–0.996)	0.102	0.283	

[a]Corrected for multiple assessments.

prespecified timepoints. This is a major strength of this study because cross-sectional validity cannot be assumed stable in a heterogenic population that may change rapidly over time with regard to the characteristics being measured. Our results confirm this assumption. For example, the relationship between the CPAx and the measurement for functional independence (FIM) increased over time, whereas the relationship between strength and the CPAx item 10 decreased over time (Table 3).

Construct validity was further investigated by comparing the CPAx-GE with the IMS, a newly developed scale for the highest level of mobility [20]. The construct "mobility" is both part of the CPAx and the IMS, which was confirmed by a high correlation between these two instruments (V3: r = 0.883, p < 0.001).

However, our study showed that in comparison to the CPAx-GE the IMS had high floor (V1: 50%) and ceiling effects (V4: 34%). Other studies also reported high floor effects (96%) at ICU admission [21]. This potentially limits the sensitivity of the IMS to detect early changes and could reduce its use to the ICU discharge, despite its recognised, excellent measurement properties.

In summary, with more than 80% of the pre-specified hypotheses accepted at baseline, critical care and hospital discharge, this study established construct, cross-sectional, and cross-cultural validity of the CPAx-GE. As a consequence, the CPAx-GE can be recommended for these three clinically relevant timepoints in critically ill patients' pathways of recovery.
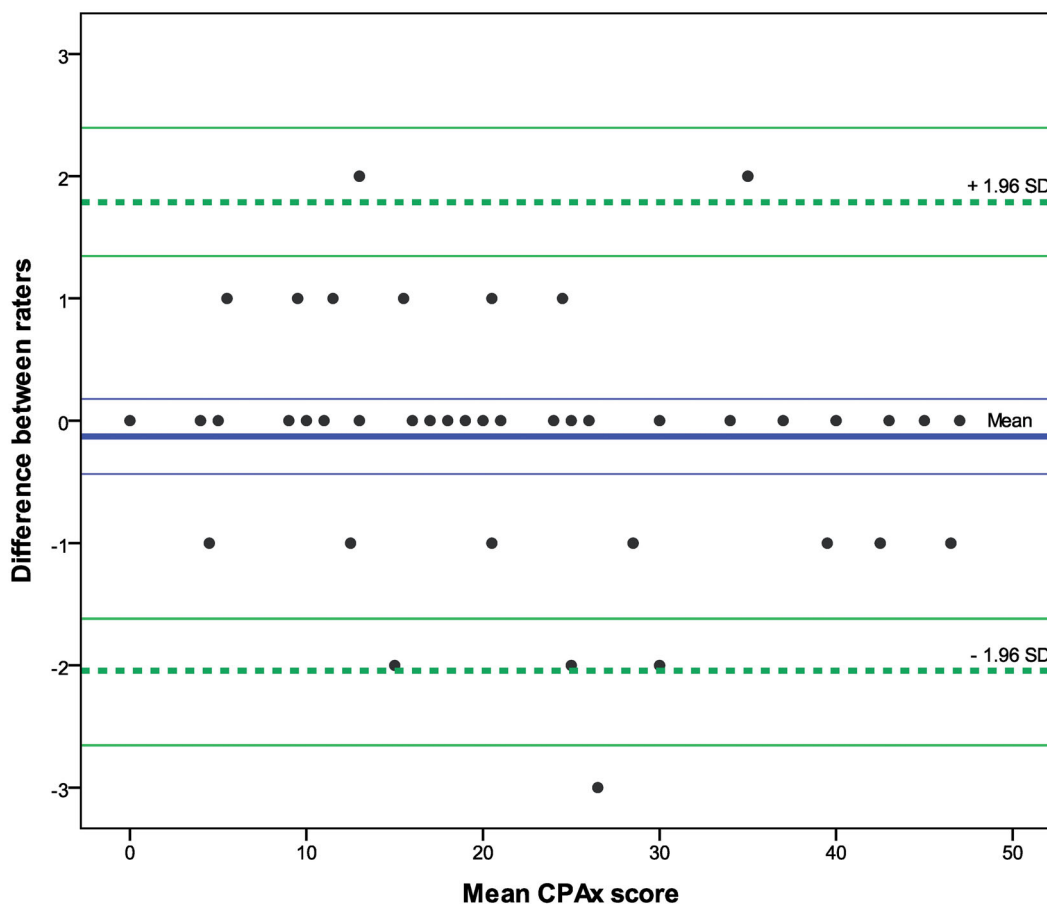
Figure 3. The modified Bland-Altman plot compares the treating versus the observing physiotherapist (n = 49) with the modified "true value varies" method which adjusts for repeated measurements with an unequal number of measurements for each individual. Mean/bias = −0.13 ± 0.15 (95% CI −0.43 to 0.18) and 95% limits of agreement: lower limit of agreement = −2.04 (95% CI −2.65 to −1.60), upper limit of agreement = 1.79 (95% CI 1.35 to 2.40), within-subject variance = 0.50 ± 0.22, between-subjects variance = 0.46 ± 0.28.

Interrater reliability during the ICU stay for the original CPAx was excellent (ICC = 0.988) [10] and has been confirmed with international raters observing two online case studies (case 1: ICC = 0.996, case 2: ICC = 0.999) by Corner et al. [36]. We found equally high ICCs (0.996) for the CPAx-GE over all four visits in clinical practice. The highest calculated SEM and SDC at ICU discharge were 0.688 and 1.89, respectively. These results are comparable to the Swedish CPAx with an ICCs of 0.97, SEM = 1.79, and SDC = 4.95 [34]. The slightly higher SDC is similar to our constructed limits of agreement. A change in the CPAx-GE of more than four points seems therefore unlikely to be attributable to measurement error. Clinical relevance may be reached in change-scores larger than six points [33].

Internal consistency of the CPAx-GE was good during the ICU stay and moderate at hospital discharge. The CPAx-GE involves respiratory and muscle functions, movements and mobility [12]. Considering that strength does not imply function, the inter-relatedness of these items may change over time. Indeed, the CPAx-GE at ICU baseline was mainly determined by respiratory function and movement, while at ICU discharge basic activities had started to emerge and at hospital discharge standing, trans-ferring, and stepping became more prevalent (Supplementary Material 4). ICU-AW is a complex, multifactorial complication [4,37] and many ICU survivors suffer from physical, cognitive, and mental impairments [38]. The moderate consistency across the CPAx-GE items might therefore be a reflexion of this multifaceted syndrome.

This study has limitations. First, all functional measurements were assessed by the treating physiotherapist and were not inde-pendent of the collection of the CPAx-GE score. This could not be avoided without increasing physiotherapy service, because most of these assessments were based on the observation of a routine physiotherapy session. In contrast, GCS, RASS, and SOFA were independently obtained, but two of the six (33%) hypotheses were rejected. The theoretical background to including these measurements was partly due to the original study and partly because mental functions are imperative to perform activities. Additionally, severity of illness is known as an important factor in the development of ICU-AW [39]. Considering our results, the rela-tionship of these factors with the CPAx might be less strong than in our theoretical model (Supplementary Material 1). Second, des-pite increasing the sample size for interrater-reliability as much as feasible, our study just barely reached the recommended size (n = 50) for all visits. Therefore, results about reliability should be interpreted with caution and measurement error might be slightly higher. Nevertheless, our results are very similar to the original and Swedish CPAx versions [10,34,36]. Additionally, it should be considered that, so far, interrater-reliability was only established by observation of the same physiotherapy treatment. Interrater-reliability on two different occasions or test–retest reliability might not be as high. Future studies are necessary, yet challenging in the ICU setting where changes may occur rapidly. Third, the limi-tations of the original CPAx (e.g., the lack of a handgrip strength protocol) also apply to the CPAx-GE. Additionally, the CPAx does not consider walking distance or exercise tolerance which should

be assessed with other instruments such as the IMS, Functional Status Score for the ICU (FSS-ICU) [40] or Physical Function ICU Test – scored (PFIT-s) [41]. Fourth, this study was conducted in a tertiary academic Swiss hospital with an advanced early rehabilitation pathway [42]. This limits generalisability to other settings with lower early mobilisation rates. Nevertheless, our results showed that both validity and reliability of the CPAx-GE were stable concepts within all investigated timepoints in the pathway of the critically ill adult regardless of its numeric score. Finally, all assessors in this study received a short training. To fully establish the CPAx-GE as a key tool in ICU rehabilitation, it would be recommendable to translate the official, freely available English e-learning into German. Future research areas include the investigation of the minimal important change, predictive validity and reliability of the CPAx-GE in more German-speaking ICU settings and areas.

Conclusions

The newly translated and cross-culturally adapted CPAx-GE was found to be conceptionally equivalent to the original. The CPAx-GE further demonstrated excellent measurement properties in critically ill adults with prolonged mechanical ventilation ($\geq$72 h) across the ICU and hospital stay. The newly established cross-sectional validity of the CPAx allows its use over three clinically relevant timepoints in critically ill patients' pathways of recovery. Therefore, the CPAx-GE appears to be a key measurement instrument for German-speaking, qualified physiotherapists to assess physical function and activity during early rehabilitation in the ICU and hospital.

ORCID

Sabrina Eggmann [iD] http://orcid.org/0000-0002-7379-7343

Data availability statement

The dataset supporting the conclusions of this article is included within the article as a Supplementary Material (Supplementary Material 5).

References

[1] Stevens RD, Marshall SA, Cornblath DR, et al. A framework for diagnosing and classifying intensive care unit-acquired weakness. Crit Care Med. 2009;37(10 Suppl.):S299–S308.

[2] De Jonghe B, Sharshar T, Lefaucheur JP, et al. Paresis acquired in the intensive care unit: a prospective multicenter study. JAMA. 2002;288(22):2859–2867.

[3] Schefold JC, Bierbrauer J, Weber-Carstens S. Intensive care unit-acquired weakness (ICUAW) and muscle wasting in critically ill patients with severe sepsis and septic shock. J Cachexia Sarcopenia Muscle. 2010;1(2):147–157.

[4] Hodgson CL, Tipping CJ. Physiotherapy management of intensive care unit-acquired weakness. J Physiother. 2017; 63(1):4–10.

[5] Fan E, Dowdy DW, Colantuoni E, et al. Physical complications in acute lung injury survivors: a two-year longitudinal prospective study. Crit Care Med. 2014;42(4):849–859.

[6] Eggmann S, Luder G, Verra ML, et al. Functional ability and quality of life in critical illness survivors with intensive care unit acquired weakness: a secondary analysis of a randomised controlled trial. PLOS One. 2020;15(3):e0229725.

[7] Zhang L, Hu W, Cai Z, et al. Early mobilization of critically ill patients in the intensive care unit: a systematic review and meta-analysis. PLoS One. 2019;14(10):e0223185.

[8] Tipping CJ, Harrold M, Holland A, et al. The effects of active mobilisation and rehabilitation in ICU on mortality and function: a systematic review. Intensive Care Med. 2017; 43(2):171–183.

[9] Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737–745.

[10] Corner EJ, Wood H, Englebretsen C, et al. The Chelsea Critical Care Physical Assessment Tool (CPAx): validation of an innovative new tool to measure physical morbidity in the general adult critical care population; an observational proof-of-concept pilot study. Physiotherapy. 2013;99(1): 33–41.

[11] Corner EJ, Soni N, Handy JM, et al. Construct validity of the Chelsea Critical Care Physical Assessment Tool: an observational study of recovery from critical illness. Crit Care. 2014; 18(2):R55.

[12] Gonzalez-Seguel F, Corner EJ, Merino-Osorio C. International Classification of Functioning, Disability, and Health domains of 60 physical functioning measurement

instruments used during the adult intensive care unit stay: a scoping review. Phys Ther. 2019;99(5):627–640.

[13] Parry SM, Granger CL, Berney S, et al. Assessment of impairment and activity limitations in the critically ill: a systematic review of measurement instruments and their clinimetric properties. Intensive Care Med. 2015;41(5):744–762.

[14] Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8(2):94–104.

[15] Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. Spine. 2000;25(24):3186–3191.

[16] Polit DF, Beck CT, Owen SV. Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. Res Nurs Health. 2007;30(4):459–467.

[17] von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370(9596):1453–1457.

[18] Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.

[19] Hermans G, Clerckx B, Vanhullebusch T, et al. Interobserver agreement of Medical Research Council sum-score and handgrip strength in the intensive care unit. Muscle Nerve. 2012;45(1):18–25.

[20] Hodgson C, Needham D, Haines K, et al. Feasibility and inter-rater reliability of the ICU Mobility Scale. Heart Lung. 2014;43(1):19–24.

[21] Tipping CJ, Bailey MJ, Bellomo R, et al. The ICU Mobility Scale has construct and predictive validity and is responsive. Ann Am Thorac Soc. 2016;13(6):887–893.

[22] Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med. 1996;22(7): 707–710.

[23] Arts DG, de Keizer NF, Vroom MB, et al. Reliability and accuracy of Sequential Organ Failure Assessment (SOFA) scoring. Crit Care Med. 2005;33(9):1988–1993.

[24] Teasdale G, Maas A, Lecky F, et al. The Glasgow Coma Scale at 40 years: standing the test of time. Lancet Neurol. 2014;13(8):844–854.

[25] Sessler CN, Gosnell MS, Grap MJ, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. Am J Respir Crit Care Med. 2002;166(10):1338–1344.

[26] de Vet HCW, Terwee CB, Mokkink LB, et al. Measurement in medicine: a practical guide. New York: Cambridge University Press; 2011.

[27] Kohn M. Sample size calculators. Clinical & Translational Science Institute UCSF; 2017 [cited 2017 Jan 5]. Available from: http://www.sample-size.net/correlation-sample-size/

[28] Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. Psychol Bull. 1979;86(2):420–428.

[29] Revelle W. psych: procedures for personality and psychological research. Evanston (IL): Northwestern University; 2020. Available from: https://CRAN.R-project.org/package=psych. Version = 2.0.12

[30] Altman DG, Bland JM. Measurement in medicine: the analysis of method comparison studies. J R Stat Soc Ser D (the Statistician). 1983;32(3):307–317.

[31] Olofsen E, Dahan A, Borsboom G, et al. Improvements in the application and reporting of advanced Bland–Altman methods of comparison. J Clin Monit Comput. 2015;29(1): 127–139.

[32] de Vet HC, Terwee CB, Knol DL, et al. When to use agreement versus reliability measures. J Clin Epidemiol. 2006; 59(10):1033–1039.

[33] Corner EJ, Hichens LV, Attrill KM, et al. The responsiveness of the Chelsea Critical Care Physical Assessment tool in measuring functional recovery in the burns critical care population: an observational study. Burns. 2015;41(2): 241–247.

[34] Holdar U, Eriksson F, Siesage K, et al. Cross-cultural adaptation and inter-rater reliability of the Swedish version of the Chelsea critical care assessment tool (CPAX-Swe) in critically ill patients. Disabil Rehabil. 2019;1–5.

[35] Astrup K, Corner EJ, Hansen MG, et al. Translation and cross-cultural adaptation of the Chelsea Critical Care Physical Assessment tool into Danish. Physiother Theory Pract. 2020;36(9):1027–1034.

[36] Corner EJ, Handy JM, Brett SJ. eLearning to facilitate the education and implementation of the Chelsea Critical Care Physical Assessment: a novel measure of function in critical illness. BMJ Open. 2016;6(4):e010614.

[37] Berger D, Bloechlinger S, von Haehling S, et al. Dysfunction of respiratory muscles in critically ill patients on the intensive care unit. J Cachexia Sarcopenia Muscle. 2016;7(4): 403–412.

[38] Needham DM, Davidson J, Cohen H, et al. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. Crit Care Med. 2012;40(2):502–509.

[39] Witteveen E, Wieske L, van der Poll T, et al. Increased early systemic inflammation in ICU-acquired weakness: a prospective observational cohort study. Crit Care Med. 2017; 45(6):972–979.

[40] Thrush A, Rozek M, Dekerlegand JL. The clinical utility of the functional status score for the intensive care unit (FSS-ICU) at a long-term acute care hospital: a prospective cohort study. Phys Ther. 2012;92(12):1536–1545.

[41] Denehy L, de Morton NA, Skinner EH, et al. A physical function test for use in the intensive care unit: validity, responsiveness, and predictive utility of the physical function ICU test (scored). Phys Ther. 2013;93(12):1636–1645.

[42] Eggmann S, Verra ML, Luder G, et al. Effects of early, combined endurance and resistance training in mechanically ventilated, critically ill patients: a randomised controlled trial. PLoS One. 2018;13(11):e0207428.