ORIGINAL ARTICLE

UDC: 616.12-008.46-036.86-07
DOI: https://doi.org/10.2298/VSP250528075I

Validity and reliability of the A-test in assessing the functional capacity of patients with heart failure

Validnost i pouzdanost A-testa u proceni funkcionalnog kapaciteta obolelih od srčane insuficijencije

Dejan Ilić*†, Dragana Tomić†, Jovana Djurdjević†, Slobodan Obradović†‡

Military Medical Academy, *Clinic for Physical Medicine and Rehabilitation, ‡Clinic for Cardiology, Belgrade, Serbia; †University of Defence, Faculty of Medicine of the Military Medical Academy, Belgrade, Serbia

Abstract

Background/Aim. Before initiating early rehabilitation, it is essential to assess the functional status of patients with heart failure (HF). Although the Barthel Index (BI) is well-established, the A-test, originally developed for traumatology and orthopedic patients, has recently been introduced into broader clinical practice. The aim of this study was to evaluate the validity and reliability of the A-test in assessing functional status in patients with HF. **Methods.** The study included patients of both sexes with HF, regardless of age or left ventricular ejection fraction, classified according to the New York Heart Association (NYHA) I–IV class, with Mini-Mental State Examination (MMSE) scores ≥ 24, who were tested daily at the same time. Two experienced senior physiotherapists performed the testing, and they completed the A-test and BI separately. **Results.** A total of 77 patients were enrolled, two-thirds of whom were male, with a mean age of 72 years and a mean ejection fraction of 41.83% [standard deviation (SD) = 11.46]. The average NYHA class at admission, 2.75 (SD = 1.05), and discharge, 2.42 (SD = 1.12), showed a statistically significant improvement (Z = -4.914, $p < 0.001$). Content validity was supported by low floor (minimal score = 2.0%) and ceiling (maximal score = 10.8%) effects, a wide range of scores (0–50), and appropriate score distribution. Concurrent validity was confirmed by a strong correlation between the A-test and BI (ϱ = 0.991, $p < 0.001$). Predictive validity was demonstrated, as the A-test score on the first day significantly predicted BI at discharge (β = 0.880, $p < 0.001$, 95% confidence interval: 1.087–1.562). Construct validity was confirmed through four hypotheses. Patients in NYHA class I–II, with higher MMSE scores and higher left ventricular ejection fractions, scored significantly better on the A-test. Age had a significant effect on the results. The A-test demonstrated excellent internal consistency (α = 0.971) and high inter-rater reliability ($\varkappa$ = 0.842). **Conclusion.** The A-test is a valid and reliable instrument for evaluating the functional capacity of patients with HF.

Key words:
heart failure; physical and rehabilitation medicine; predictive value of tests; rehabilitation.

Apstrakt

Uvod/Cilj. Pre početka rane rehabilitacije, potrebno je obolelima od srčane insuficijencije (*heart failure* – HF) uraditi procenu funkcionalnog statusa. Iako je Bartelov indeks (BI) dobro poznat, nedavno je u širu kliničku praksu uveden A-test, osmišljen primarno za pacijente iz oblasti traumatologije i ortopedije. Cilj studije bio je da se ispita validnost i pouzdanost A-testa za procenu funkcionalnog stanja obolelih od HF. **Metode.** U studiju su bili uključeni oboleli od HF, pripadnici oba pola, bez obzira na starost ili ejakcionu frakciju leve komore, klasifikovani prema *New York Heart Association* (NYHA) I–IV klase, sa rezultatima ≥ 24 na mini-testu mentalnog stanja (*Mini-Mental State Examination* – MMSE), koji su testirani svakog dana u isto vreme. Dva iskusna viša fizioterapeuta su vršila testiranje i odvojeno popunjavala A-test i BI. **Rezultati.** Studijom je obuhvaćeno ukupno 77 bolesnika, dve trećine muškog pola, prosečne starosti 72 godine i prosečne ejekcione frakcije 41,83% [standardna devijacija (SD) = 11,46]. Prosečna NYHA klasa na prijemu iznosila je 2,75 (SD = 1,05) a na otpustu 2,42 (SD = 1,12), što je predstavljalo statistički značajno poboljšanje (Z = -4,914, $p < 0,001$). Validnost sadržaja potvrđena je niskim efektima poda (minimalni skor = 2,0%) i plafona (maksimalni skor = 10,8%), širokim opsegom vrednosti skorova (0–50) i odgovarajućom distribucijom skorova. Konkurentna validnost je potvrđena jakom korelacijom između A-testa i BI (ϱ = 0,991, $p < 0,001$). Prediktivna validnost je dokazana jer je skor A-testa prvog dana značajan prediktor za BI na otpustu

Correspondence to: Slobodan Obradović, Military Medical Academy, Clinic for Cardiology, Crnotravska 17, 11 000 Belgrade, Serbia. E-mail: sloba.d.obradovic@gmail.com

(β = 0,880, p < 0,001, 95% *confidence interval*: 1,087–1,562). Konstruktivna validnost potvrđena je kroz četiri hipoteze. Bolesnici NYHA I i II klase, sa višim rezultatima na MMSE testu i višim ejekcionim frakcijama leve komore, pokazali su značajno bolje rezultate na A-testu. Godine starosti imale su značajan uticaj na rezultate. A-test je pokazao izuzetnu internu konzistentnost (α = 0,971) i visoku pouzdanost između ocenjivača ($\varkappa$ = 0,842). **Zaključak.** A-test je validan i pouzdan instrument za procenu funkcionalne sposobnosti obolelih od HF.

Ključne reči:
srce, insuficijencija; medicina, fizikalna i rehabilitacija; testovi, prognostička vrednost; rehabilitacija.

Introduction

Heart failure (HF) is the inability of the heart to pump sufficient amounts of blood to peripheral tissues and is the leading cause of death in the developed world [1–3]. More than 64 million people suffer from HF, half of whom have reduced left ventricular ejection fraction (EF) – LVEF [1, 2, 4]. The five-year mortality rate is 50–75% [2]. Almost all heart diseases lead to HF, most commonly arterial hypertension [5], atrial fibrillation [6], anemia [7], as well as endocrine diseases, primarily diabetes [8, 9]. Symptoms of HF include fatigue, shortness of breath, and inability to sleep on a flat surface. Cases of HF with preserved EF, intermediate range, and reduced LVEF have been described. Preserved LVEF, ≥ 50%, is indicative of diastolic HF, moderately preserved LVEF represents a "gray zone" for patients with values between 40–49%, whereas those with LVEF < 40% are classified as having reduced LVEF [10].

According to the New York Heart Association (NYHA), the NYHA classification is based on the severity of symptoms and physical activity, dividing patients into four groups: Class I – the patient has no limitations in physical activity; Class II – usual activities cause fatigue, shortness of breath, or palpitations; Class III – the patient has significant limitations in physical activity; Class IV – symptoms of HF are present at rest and worsen with minimal physical exertion [10].

According to our study results, cardiac rehabilitation and acupuncture, in addition to drug therapy, may be beneficial in improving the functional ability of HF patients as measured by the A-test [11]. Exercise is the gold standard in the rehabilitation of cardiac patients. Before starting rehabilitation, it is necessary to assess the functional status of patients and continue to monitor their recovery.

The Barthel Index (BI) is commonly used in the rehabilitation of cardiac patients with HF [12–14]. BI assesses the functional independence of patients (feeding, dressing, toileting, personal hygiene, transfers, walking, bowel and bladder control), mainly in the self-care and walking and moving domain according to the International Classification of Functioning, Disability, and Health (ICF) established by the World Health Organization in 2001. A BI score < 85 at discharge indicates increased mortality [15]. BI has a maximum value of 100 when the patient is independent, and a minimum value of 0, indicating complete dependence.

The A-test, created in the early 2000s for evaluating functional recovery, was primarily used to assess patients during the early stages of rehabilitation in orthopedic wards. Today, however, it serves as a useful assessment tool in the early rehabilitation of patients who have undergone cardiac, thoracic, vascular, or neurosurgical procedures [16–18]. Similar to the BI, it tests ten patient activities, mainly within the domain of changing basic body positions and walking and moving according to the ICF, using a 6-level ordinal scale (0–5), where 0 represents the minimum and 50 the maximum score. The A-test is a reliable instrument for everyday evaluation of functional recovery during early rehabilitation of patients surgically treated in an orthopedic ward [16]. The A-test can identify which patients are progressing and which are lagging behind in functional recovery on each day of rehabilitation [17]. Patients should not be assessed according to diagnosis, but rather according to the type and severity of functional deficit [18–20].

The aim of this study was to evaluate the validity and reliability of the A-test in patients with HF. Validity refers to the extent to which an instrument measures what it is intended to measure [18, 21]. Reliability refers to the ability of an instrument to produce consistent results when applied two or more times [21].

Methods

This was a prospective, clinically controlled study, with one experimental group of 77 patients tested using two functional tests by two examiners. The research was conducted at the Clinic for Cardiology of the Military Medical Academy (MMA), Belgrade, Serbia. The research team consisted of a cardiologist, a physiatrist, a speech therapist trained in psychological testing, and two senior physiotherapists. The study was approved by the Ethics Committee of the MMA (No. 56/2022, from December 1, 2022) and the Ethics Committee of the Faculty of Medicine of the MMA (No. 1/4/2023, from February 21, 2023) within the framework of the doctoral dissertation decision of the Senate of the University of Defence (No. 27-272, from July 25, 2023). All patients provided written informed consent to participate in the study.

Patient selection

Patients diagnosed with HF were selected based on their clinical presentation, medical history, and echocardiographic findings. As a prerequisite for participation in the clinical study, the patient's cognitive status was determined using the Mini-Mental State Examination (MMSE), and patients with a score of 24 or above were included in the study. Patients of both sexes, without restrictions regarding age or LVEF values, classified as NYHA class I–IV, as determined by a physician at the beginning and end of hospitalization, were included.

Exclusion criteria

Immobile patients (those with quadriplegia, paraplegia, hemiplegia, or severe degenerative and inflammatory rheumatism), as well as patients with anemia (hemoglobin < 90 g/L) or other severe comorbidities that prevent functional testing, were excluded from the study. Exclusion criteria also included instability of vital parameters: arterial blood pressure > 180/120 mmHg or < 90/60 mmHg, heart rate ≥ 130 beats *per* min, malignant cardiac arrhythmias, febrile conditions, or instability of vital signs within the previous 24 hrs.

A-test and Barthel Index

Two experienced senior physiotherapists tested each patient together at the same time every day during hospitalization, and filled out the tests separately. The A-test evaluates ten activities: changing position in bed (from supine to side-lying and from supine to sitting on the edge of the bed), getting in and out of bed, standing, walking, sitting in a chair, and using the toilet. The last two activities are walking up and down the stairs and walking endurance. The A-test uses a 6-level ordinal scale ranging from 0 to 5, where a score of 0 indicates that the patient is unable to perform the activity, and a score of 5 represents independent and safe performance (Appendix 1). Walking endurance is scored separately: 0 for inability to walk; 1 for walking up to 5 m (which is room-level walking); 2 for walking up to 15 m; 3 for walking up to 50 m; 4 for walking up to 100 m; and 5 for walking over 100 m. The minimum A-test score is 0, while the maximum is 50 [16]. The well-known BI also assesses ten activities, using different scales (from a 2-level ordinal scale for bathing to a 4-level ordinal scale for mobility on a level surface), with 0 as the minimum and 100 as the maximum score.

Feeding, personal toileting, bathing, dressing, transfers when using the toilet and bed-chair, walking and stair climbing, and sphincter control were tested [22, 23].

Sample size and study power

The sample size required to test the validity and reliability of the A-test in patients with HF across all four NYHA classes was a minimum of 60 subjects. The sample size of 60 subjects was determined according to the standard formula for repeated measures analysis of variance, assuming an alpha (α) error of 0.05, a study power of 80% (β = 0.20), and an internal measurement correlation of ρ = 0.50. The sample size (N) was calculated according to the formula: $N = [2 \times (Z_{1-\alpha/2} + Z_{1-\beta})^2 \times (1 - \rho)] / (m \times f^2)$, i.e., $N = [2 \times (1.96 + 0.84)^2 \times (1–0.50)] / (m \times f^2) = 60$.

Statistical analysis

The validity of the A-test was assessed through three complementary aspects of validity: content, criterion validity, and construct validity. Content validity was determined by analyzing the distribution of results, floor and ceiling effects, and the width of the range of results. The following statistical indicators were used: the percentage of minimum and maximum results by day (floor/ceiling effect), skewness (Sk) values, and the theoretical range of the scale (0–50). Criterion validity was assessed in two ways: concurrent validity was examined using Spearman's correlation between A-test results and BI; predictive validity was examined using univariate linear regression analysis, which found that A-test results on day 1 significantly predicted functional outcome (BI on day 6). Construct validity was assessed by testing hypothetically expected differences in test results in accordance with theoretical assumptions. This validity was measured by the Mann-Whitney (*U*) test and Spearman's product-moment correlation coefficient. Reliability of the A-test was assessed in two ways: internal consistency and inter-rater reliability. Cronbach's alpha coefficient was used to assess internal reliability. The agreement between two independent raters was examined using the kappa (κ) coefficient, which measures inter-rater reliability. Intraclass Correlation Coefficient (ICC) was used to evaluate the A-test's stability across time points. The ICC was calculated using a two-way random effects model with absolute agreement and average measures – ICC (2, k). The normality of distribution for continuous variables was assessed using the Shapiro-Wilk test. Categorical variables were summarized using frequencies and percentages. For continuous variables, those following a normal distribution were presented as means with standard deviations (SD), while non-normally distributed variables were described using medians (Me). Statistical analysis was performed using the IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, NY, USA).

Results

The study sample consisted of a total of 77 patients, the majority of whom were male, i.e., 66.2%, while 33.8% were female. The majority of participants were older than 65 years (80.5%), with a mean age of 72.04 years (SD = 11.12), a median of 72 years, and a range of 38 to 99 years. Cognitive functioning, assessed by MMSE, showed that the subjects scored an average of 27.42 points (SD = 2.14), indicating preserved cognition. The median MMSE score was 28 points, with a range of 24 to 30. Regarding cardiac function, the mean EF was 41.83% (SD = 11.46), with a median of 44% and values ranging from 20% to 66%. Functional status was also assessed using the NYHA classification. At admission, the mean NYHA class was 2.75 (SD = 1.05), while a slight improvement was noted at discharge with a mean value of 2.42 (SD = 1.12). The range of the NYHA class ranged from I to IV in both measurements. The Wilcoxon test indicates a significant difference in NYHA scores between admission and discharge ($Z = -4.914$, $p < 0.001$). At admission, 15.6% of patients were classified as NYHA class I, 23.4% as NYHA class II, 31.2% as NYHA class III, while 29.9% of patients were in the NYHA class IV. At discharge, 28.6% of patients were in the NYHA class I, 22.1% in NYHA class II, 28.6% in NYHA class III, while 20.8% of patients remained in the NYHA class IV (Table 1). The sum of the total num-

ber of measurements *per* therapist was 444, which means the total number of measurements taken by two therapists was 888. The smallest number of measurements *per* patient was 3, while the largest was 13 for each therapist.

Descriptive statistics for the results of the A-test and BI during the 13 days of hospitalization are presented. Based on the values, a gradual increase in the average results of both tests can be observed during the first ten days, which indicates progress in the functional recovery of the patients. After the tenth day, stable values are recorded, with reduced variability, which may be a consequence of the reduced number of subjects. The total median values for the entire period were Me = 36.36 for the A-test and Me = 73.64 for BI (Table 2). The A-test was validated based on three aspects of validity: content validity, criterion validity, and construct validity.

Table 1

General data about patients

Parameter	Values
Gender	
male	51 (66.2)
female	26 (33.8)
Age, years	
≤ 65	15 (19.5)
> 65	62 (80.5)
mean ± SD; median (range)	72.04 ± 11.12; 72 (38–99)
MMSE, mean ± SD; median (range)	27.42 ± 2.14; 28 (24–30)
Ejection fraction, mean ± SD; median (range)	41.83 ± 11.46; 44 (20–66)
NYHA categories at admission	
I	12 (15.6)
II	18 (23.4)
III	24 (31.2)
IV	23 (29.9)
NYHA at admission*, mean ± SD; median (range)	2.75 ± 1.05; 3 (1–4)
NYHA at discharge*, mean ± SD; median (range)	2.42 ± 1.12; 2 (1–4)
NYHA categories at discharge	
I	22 (28.6)
II	17 (22.1)
III	22 (28.6)
IV	16 (20.8)

MMSE – Mini-Mental State Examination; SD – standard deviation; NYHA – New York Heart Association.
Data are shown as numbers (percentages), or mean ± SD; median (range), where indicated.
***Note*. *The Wilcoxon test showed a statistically significant difference in NYHA scores between admission and discharge ($p < 0.001$).**

Table 2

Minimum, maximum, and median values of the A-test and Barthel Index

Parameter	Minimum	Maximum	Median
A-test			
day 1	0	50	36
day 2	0	50	36
day 3	0	50	37
day 4	0	50	37
day 5	0	50	38
day 6	3	50	41
day 7	0	50	40
day 8	0	45	39.5
day 9	5.5	43	36
day 10	29	43	40.5
day 11	33	42	39
day 12	33	33	33
day 13	33	33	33
overall	1	50	36.36
Barthel Index			
day 1	0	100	65
day 2	0	100	70
day 3	0	100	70
day 4	5	100	75

Table 2 (continued)

Parameter	Minimum	Maximum	Median
day 5	5	100	75
day 6	10	100	80
day 7	10	100	80
day 8	5	70	61
day 9	20	90	67.5
day 10	55	90	87.5
day 11	55	90	85
day 12	55	55	55
day 13	55	55	55
overall	7.5	100	73.64

Table 3

Floor and ceiling effect for the A-test

Score (%)	Days of rehabilitation												
	1	2	3	4	5	6	7	8	9	10	11	12	13
Minimal	3.9	3.9	2.6	1.5	1.9	0	4.0	8.3	0	0	0	0	0
Maximal	18.2	18.2	19.5	16.9	18.9	28.2	20.0	0	0	0	0	0	0

Table 4

Distribution characteristics of the A-test by day and overall

Parameter	Day													Overall
	1	2	3	4	5	6	7	8	9	10	11	12	13	
n	77	77	77	65	53	39	25	12	8	6	3	1	1	77
Sk	-0.322	-0.457	-0.740	-0.883	-0.961	-1.373	-1.386	-1.789	-1.570	-0.923	-0.935	/	/	-0.790
Ku	-1.308	-1.080	-0.526	-0.098	-0.009	1.583	1.218	2.083	2.623	-0.930	/	/	/	-0.324

n – number of patients; Sk – skewness; Ku – kurtosis.

***Note*: The symbol "/" indicates that the correlation for days 11, 12, and 13 of follow-up was not tested due to the small number of subjects.**

Content validity

We assumed that the A-test would demonstrate a satisfactory level of content validity if the scores obtained during the 13 days of hospitalization covered the full range from 0 to 50, if the skewness value of the distribution was below 1.00, and if fewer than 15% of the scores represented the lowest or highest possible value on the scale. To determine content validity, we used the floor and ceiling effect, measures of asymmetry, and the range of values.

The floor and ceiling effects were analyzed throughout the 13 days of early rehabilitation. Content validity, which assesses the extent to which the instrument covers key aspects of functional capacity, was confirmed by the wide score range, low floor (minimal score 2.0%) and ceiling (maximal score 10.8%) effects, and symmetrical distribution (Table 3).

The distribution shape parameters (skewness and kurtosis) were analyzed across all 13 days of early rehabilitation, with the number of respondents gradually decreasing, particularly affecting the reliability of estimates after day 8 ($N < 10$). From day 1 to day 10, all scores showed negative skewness, suggesting that the majority of participants achieved higher functional scores. Notably, from day 6 to 9, skewness values exceeded ±1, indicating a significant deviation from symmetry (Table 4).

Kurtosis values were negative from day 1 to 5, reflecting flatter-than-normal distributions, while a shift to positive values occurred from day 6 to day 9, suggesting increased score concentration around the mean and reflecting a more peaked (leptokurtic) distribution during this later phase of rehabilitation (e.g., day 6 = 1.583; day 9 = 2.623) (Figure 1).

When considering the total A-test scores across all 13 days, the average skewness (Sk = -0.790) and kurtosis (Ku = -0.324) remained within acceptable limits (±1), indicating an approximately symmetric and moderately flattened distribution. These findings suggest that the test was capable of capturing variations in functional capacity across time without substantial ceiling or floor effects, fulfilling the second criterion of content validity (Figure 1).

The third criterion of content validity, which relates to the range of values, was also met, given that the range of values ranged from 0 to 50, which is both the theoretical minimum and maximum.

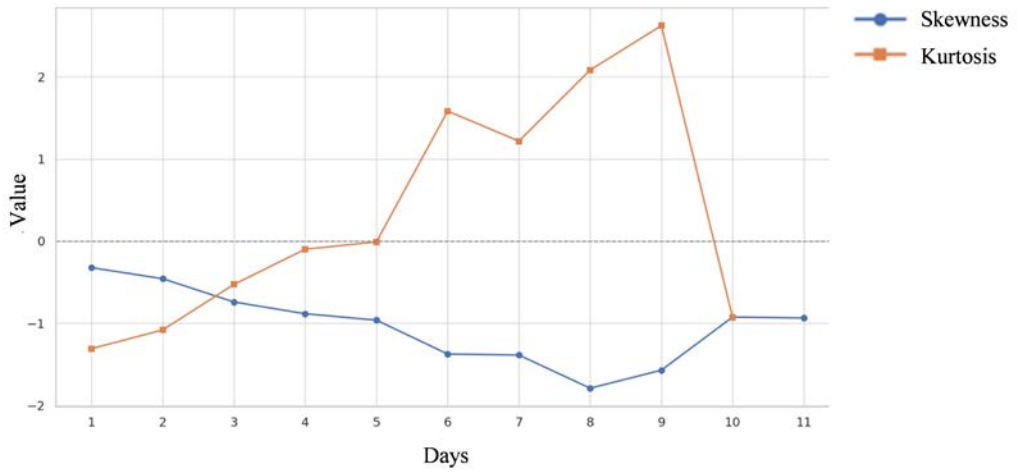

Fig. 1 – Skewness and kurtosis of A-test scores by day.

Table 5

Spearman's correlation between the A-test composite scores and the Barthel Index

Parameter	Value
ρ	0.991
p-value	< 0.001
n	77

ρ – Spearman's correlation coefficient; *p* – statistical significance; n – number of patients.

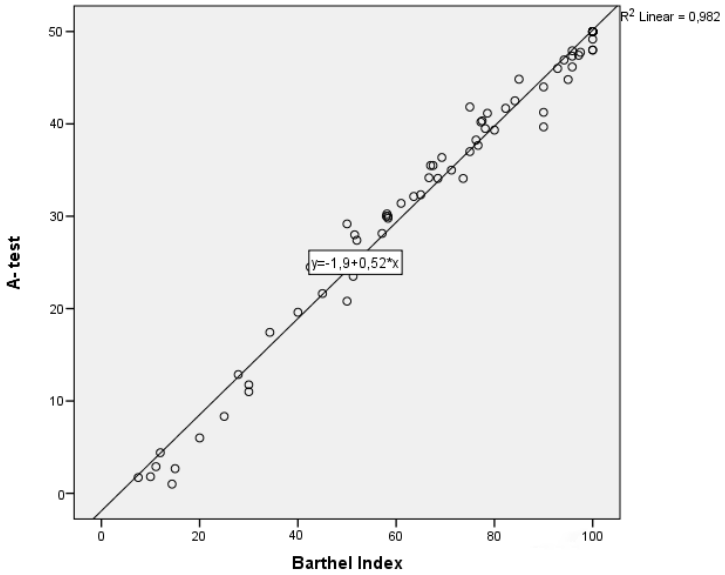

Fig. 2 – Correlation between A-test composite scores and Barthel Index.

Criterion validity

Criterion validity refers to the assessment of whether a test adequately predicts or correlates with external criteria. In this context, we examined two types of validity: concurrent validity and predictive validity. Since the A-test aims to assess the functional status of patients, the results of BI were used for the criterion assessment of concurrent validity. Spearman's correlation coefficient (rho) was used to test association.

Using Spearman's correlation coefficient, we examined whether the results of the A-test were statistically significantly related to BI. The results show a very high correlation ($\rho = 0.991$, $p < 0.001$) (Table 5). Such a high correlation indicates almost identical results of the two tests (Figure 2).

For testing the correlation, the averages of both researchers and the average for all testing days were taken.

The correlation between the results of the A-test and BI over a ten-day period was determined, using Spearman's product-moment correlation coefficient. The results show a very strong positive and statistically significant correlation between the results of these two tests over all ten days. The correlations range from ρ = 0.929 (day 9) to ρ = 0.989 (day 2) (Table 6).

The predictive validity of the A-test was examined through the ability of the test to predict the later functional outcome of patients. For this purpose, univariate linear regression analysis was used to examine whether the results of the A-test on day 1 were a statistically significant predictor of BI at discharge. In order to ensure a sufficient number of subjects, day 6 was taken as the date of discharge.

The analysis of the prediction of BI at discharge using the initial A-test score as a predictor is presented. The data indicate that the A-test measured on the first day is a significant predictor of BI at discharge [β = 0.880, $p < 0.001$, 95% confidence interval (CI): 1.087–1.562]. The coefficient of determination is very high, and the value $R^2 = 0.769$ means that the model explains 76.9% of the variance of BI at discharge, which indicates a very good predictive power of this model (Table 7).

Construct validity

The results of the A-test are presented according to the age of the respondents. The average results of the A-test were better in respondents younger than 65 years (Me = 48.00) compared to those older than 65 years (Me = 32.24). This difference was statistically significant ($U = 135.00$, $p < 0.001$) (Table 8). The first hypothesis, that the results of the A-test would be statistically significantly better in those younger than 65 years compared to those older than 65 years, was confirmed.

Table 9 shows the results of the A-test according to the NYHA class at admission and at discharge. At admission, patients in the NYHA class I had a median A-test result of Me = 50.00, in the NYHA class II, Me = 46.09, in class III, Me = 34.59, while in the NYHA class IV, the lowest median value was recorded, Me = 17.43. The difference in A-test results between categories at admission is statistically significant ($U = 62.05$, $p = 0.000$). At discharge, the median A-test for the NYHA class I was Me = 50.00, for NYHA class II Me = 40.20, for NYHA class III Me = 30.13, while in the NYHA class IV, the lowest value was recorded, Me = 9.67. The difference in the A-test scores in relation to the NYHA class at discharge is also statistically significant ($U = 65.77$, $p = 0.000$). These findings confirm that a higher NYHA

Table 6

Correlation between the A-test and the Barthel Index by measurement days

Assessment day	ρ	*p*-value
Day 1	0.989	< 0.001
Day 2	0.989	< 0.001
Day 3	0.984	< 0.001
Day 4	0.973	< 0.001
Day 5	0.953	< 0.001
Day 6	0.976	< 0.001
Day 7	0.949	< 0.001
Day 8	0.935	< 0.001
Day 9	0.929	0.001
Day 10	0.985	< 0.001

ρ – Spearman's correlation coefficient; *p* – statistical significance.

Table 7

Prediction of the Barthel Index at discharge based on day 1 A-test score

β	*t*-value	*p*-value	95% CI limit (lower–upper)	R^2
0.880	11.298	< 0.001	1.087–1.562	0.769

CI – confidence interval.
***Note*. Univariate linear regression analysis was performed.**

Table 8

Results of the A-test according to the patient's age

Statistic	Age (years)		All patients
	≤ 65	> 65	
Minimum	34.1	1	1
Maximum	50	50	50
Median	48	32.24	36.36
U			135
p-value			< 0.001

***U* – Mann-Whitney test; *p* – statistical significance.**

class, indicating a more severe functional class, consistently accompanies lower A-test scores, both at admission and at discharge. This was our second hypothesis, which was therefore confirmed.

The correlation between the A-test results and the MMSE scale was examined. The analysis revealed a statistically significant positive correlation of medium intensity ($\rho = 0.596$, $p < 0.001$), indicating that better cognitive status, measured by the MMSE scale, was associated with better A-test results. A total of 77 subjects were included in the analysis (Table 10). Hypothesis 3, that patients with higher MMSE scores would achieve better A-test results, was thus confirmed.

The relationship between the A-test results and EF was also examined. A statistically significant positive correlation of weaker intensity was obtained ($\rho = 0.269$, $p = 0.018$), indicating that higher values of EF are associated with better results on the initial measurement of the A-test. A total of 77 subjects were included in the analysis (Table 11). Hypothesis 4, that patients with a higher EF would have better A-test results at the initial measurement, was confirmed.

Reliability of the A-test

The reliability of the A-test was measured through internal consistency and reproducibility.

Internal consistency is a form of reliability that refers to the degree to which the items of an instrument are interconnected, that is, to what extent they jointly measure the same construct. In this study, Cronbach's alpha was used as the main indicator of the internal reliability of the A-test.

The A-test demonstrated excellent internal consistency during the first 11 days of rehabilitation, with Cronbach's alpha values consistently high ($\alpha = 0.970–0.972$). The overall reliability ($\alpha = 0.971$) confirms that the test items reliably assess the same construct, functional status, throughout the rehabilitation process. These findings support the use of the A-test as a stable and consistent tool for monitoring patient progress in clinical settings (Table 12).

Reproducibility of the A-test was measured by inter-rater reliability. The A-test showed very high inter-rater reliability throughout the rehabilitation period, with κ coefficients ranging from 0.860 to 1.000. According to the Landis

Table 9

The A-test results according to the NYHA at admission and discharge

Statistic	NYHA at admission				NYHA at discharge			
	I	II	III	IV	I	II	III	IV
Minimum	46.92	35.5	20.8	1	41.15	30	20.8	1
Maximum	50	50	42.5	34.1	50	48	41.83	29.78
Median	50	46.09	34.59	17.43	50	40.2	30.13	9.67
U	62.05				65.77			
p-value	< 0.001				< 0.001			

NYHA – New York Heart Association; *U* – Mann-Whitney test; *p* – statistical significance.

Table 10

Spearman's correlation between the A-test and MMSE test scores

Parameter	Value
ρ	0.596
p-value	< 0.001
n	77

MMSE – Mini-Mental State Examination; ρ – Spearman's correlation coefficient; *p* – statistical significance; n – number of patients.

Table 11

Spearman's correlation between the A-test results at initial measurement and the ejection fraction

Parameter	Value
ρ	0.269
p-value	0.018
n	77

ρ – Spearman's correlation coefficient; *p* – statistical significance; n – number of patients.

Table 12
Reliability of the A-test

A-test	α
Day 1	0.972
Day 2	0.969
Day 3	0.970
Day 4	0.967
Day 5	0.967
Day 6	0.968
Day 7	0.968
Day 8	0.968
Day 9	0.970
Day 10	0.970
Day 11	0.970
Overall	0.971

α – Cronbach alpha coefficient.

Table 13
Inter-examiner agreement of results on the A-test

A-test	κ
Day 1	0.931
Day 2	0.945
Day 3	0.930
Day 4	0.951
Day 5	0.980
Day 6	1.000
Day 7	0.955
Day 8	0.908
Day 9	0.860
Day 10	1.000
Day 11	1.000
Day 12	n/a
Day 13	n/a
Overall	0.842

n/a – not applicable.

and Koch classification [24], all values reflect very good agreement ($\kappa \geq 0.81$), even at the lowest point ($\kappa = 0.860$ on day 9). The overall inter-rater reliability ($\kappa = 0.842$) confirms that the A-test yields consistent results across different examiners, supporting its applicability in clinical practice for reliable patient assessment (Table 13).

Discussion

The mean age of the subjects and the distribution by gender are similar to our previous studies in this area, as well as to the works we cited [11]. A statistically significant decrease in NYHA class at discharge, compared to admission, is an indicator of good, modern drug therapy for HF [25]. The A-test is designed to assess the basic functional abilities that the patient should regain during the initial phase of recovery.

Content validity refers to the extent to which an instrument covers items relevant and essential to the concept being measured. When an instrument covers all important aspects of the target area and is tailored to the purpose for which it is used, it is considered to have good content validity [26]. The A-test demonstrated good content validity, with a wide range of scores, minimal asymmetry, and a low percentage of extreme scores – indicating a low floor and ceiling effect.

Concurrent validity refers to the degree to which the A-test scores correlate with an existing BI used for the same or similar construct. If a new test shows a high correlation with existing measures, it is considered to have good concurrent validity. This type of validity is particularly important for determining whether the A-test truly measures the same thing as other validated instruments. The results indicate a stable and strong relationship between A-test scores and functional independence measured by BI in subjects across all follow-up days, with the exception that correlations for days 11, 12, and 13 were not tested due to the small number of participants.

Predictive validity refers to the ability of the A-test to predict future behavior or outcomes based on measurements. High predictive validity indicates that the test has the ability to forecast future events based on current results successfully. The A-test score on the first day is a significant predictor of BI at discharge, according to the values of the coefficient of determination.

The construct validity of the A-test has been confirmed with several measurements, where the value of the A-test depends on the patient's functional ability and age.

Reliability is a basic psychometric property of the instrument and refers to the degree to which the measurement is consistent, i.e., in which the instrument provides stable and reproducible results [21]. Internal consistency is an indicator of the homogeneity of the scale and is important for assessing whether all items of the instrument can be considered parts of the same measurement. The most commonly used indicator of internal consistency is Cronbach's alpha, which is calculated based on the average correlation between items. The Cronbach's alpha value ranges from 0 to 1, with higher values indicating greater internal consistency. It is common for $\alpha \geq 0.70$ to be considered an acceptable level of reliability for research purposes, while values above 0.80 are interpreted as good, and above 0.90 as a very high level of consistency. Internal consistency can also be used to analyze what happens to alpha if individual items are removed from the scale. This analysis helps identify items that may be disrupting the internal structure of the instrument. In our study, the A-test showed a very high level of consistency.

Inter-rater reliability refers to the degree of agreement between different raters who apply the same instrument under the same conditions. A high level of inter-rater reliability indicates that the results do not depend on who conducts the measurement, but that the instrument itself consistently measures what is being assessed. Assessments were independently recorded by two raters on the A-test for each respondent at all measurement intervals. We measured the agreement between their assessments using the κ coefficient. The usual classification of κ coefficient of agreement values is as follows: $\kappa < 0.00$ – poor agreement; $\kappa = 0.00–0.20$ – slight agreement; $\kappa = 0.21–0.40$ – fair agreement; $\kappa = 0.41–0.60$ – moderate agreement; $\kappa = 0.61–0.80$ – substantial agreement; and $\kappa = 0.81–1.00$ – almost perfect agreement. In our study, the two senior physiotherapists showed very good agreement.

The results obtained in this study show that the A-test has extremely good psychometric characteristics in terms of reliability. Both internal consistency and inter-rater reliability show very high values. Cronbach's alpha coefficients over the eleven days of measurement, as well as for the total score, are consistently above 0.97, indicating exceptional test homogeneity and consistent measurement of the same construct – the functional status of patients. At the same time, κ coefficient values, ranging from 0.860 to 1.000, indicate almost perfect agreement between the two raters. These findings confirm that the A-test is a reliable instrument for assessing the functional recovery of patients during early rehabilitation, regardless of who performs the assessment.

ICC was calculated to evaluate the temporal stability of the A-test scores across repeated measurements. The obtained ICC (2, k) value was 0.971, indicating excellent reliability and minimal within-subject variability over time.

Conclusion

The A-test demonstrated strong validity and reliability metrics, confirming its clinical utility in assessing functional capacity in patients with heart failure. Its structured scoring system allows for precise monitoring of early rehabilitation progress. The A-test is easily applied in clinical practice and monitors all patient activities from lying in bed to walking. Future studies may further explore its responsiveness and predictive power in larger and more diverse cardiac populations.

REFERENCES

1. *Shams P, Malik A, Chhabra L.* Heart Failure (Congestive Heart Failure). In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025.
2. *Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS.* Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res 2023; 118(17): 3272–87. Erratum in: Cardiovasc Res 2023; 119(6): 1453.
3. *Malik A, Brito D, Vaqar S, Chhabra L.* Congestive Heart Failure. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024.
4. *Chamberlain AM, Boyd CM, Manemann SM, Dunlay SM, Gerber Y, Killian JM,* et al. Risk Factors for Heart Failure in the Community: Differences by Age and Ejection Fraction. Am J Med 2020; 133(6): e237–48.
5. *Triposkiadis F, Sarafidis P, Briasoulis A, Magouliotis DE, Athanasiou T, Skoularigis J,* et al. Hypertensive Heart Failure. J Clin Med 2023; 12(15): 5090.
6. *Bergau L, Bengel P, Sciacca V, Fink T, Sohns C, Sommer P.* Atrial Fibrillation and Heart Failure. J Clin Med 2022; 11(9): 2510.
7. *Sabah ZU, Aziz S, Wani JI, Masswary A, Wani SJ.* The association of anemia as a risk of heart failure. J Family Med Prim Care 2020; 9(2): 839–43.
8. *Palazzuoli A, Iacoviello M.* Diabetes leading to heart failure and heart failure leading to diabetes: epidemiological and clinical evidence. Heart Fail Rev 2023; 28(3): 585–96.
9. *Siao WZ, Chen YH, Tsai CF, Lee CM, Jong GP.* Diabetes Mellitus and Heart Failure. J Pers Med 2022; 12(10): 1698.
10. *Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS,* et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC)Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016; 37(27): 2129–200. Erratum in: Eur Heart J 2018; 39(10): 860.
11. *Ilic D, Jovic Z, Mladenovic Z, Pejovic V, Lung B, Kozic A, Obradovic S.* Acupuncture and Kinesitherapy Improve Physical Activity More than Kinesitherapy Alone in Patients with Acute Decompensated Chronic Heart Failure with Reduced Ejection Fraction Who Are Already on Optimal Drug Therapy: A Randomized, Sham-Controlled, Double-Blind Clinical Study. Biomedicines 2025; 13(1): 176.
12. *Delgado B, Novo A, Lopes I, Rebelo C, Almeida C, Pestana S,* et al. The effects of early rehabilitation on functional exercise tolerance in decompensated heart failure patients: Results of a multicenter randomized controlled trial (ERIC-HF study). Clin Rehabil 2022; 36(6): 813–21.
13. *Aimo A, Barison A, Mammini C, Emdin M.* The Barthel Index in elderly acute heart failure patients. Frailty matters. Int J Cardiol 2018; 254: 240–1.
14. *Motoki H, Nishimura M, Kanai M, Kimura K.* Impact of inpatient cardiac rehabilitation on Barthel Index score and prognosis in patients with acute decompensated heart failure. Int J Cardiol 2019; 293: 125–30.

15. *Katano S, Yano T, Ohori K, Kouzu H, Nagaoka R, Honma S,* et al. Barthel Index Score Predicts Mortality in Elderly Heart Failure - A Goal of Comprehensive Cardiac Rehabilitation. Circ J 2021; 86(1): 70–8.
16. *Vukomanović A, Djurović A, Popović A, Ilić D.* The A-test – reliability of functional recovery assessment during early rehabilitation of patients in an orthopedic ward. Vojnosanit Pregl 2014; 71(7): 639–45.
17. *Vukomanovic A, Djurovic A, Brdareski Z.* Diagnostic accuracy of the A-test and cutoff points for assessing outcomes and planning acute and post-acute rehabilitation of patients surgically treated for hip fractures and osteoarthritis. Vojnosanit Pregl 2016; 73(12): 1139–48.
18. *Vukomanović A, Djurović A, Popović Z, Pejović V.* The A-test: assessment of functional recovery during early rehabilitation of patients in an orthopedic ward – content, criterion and construct validity. Vojnosanit Pregl 2014; 71(8): 715–22.
19. *Leonardi M, Lee H, Kostanjsek N, Fornari A, Raggi A, Martinuzzi A,* et al. 20 Years of ICF-International Classification of Functioning, Disability and Health: Uses and Applications around the World. Int J Environ Res Public Health 2022; 19(18): 11321.
20. *Krzoska C.* International classification of functioning, disability and health (ICF) in daily clinical practice: Structure, benefits and limitations. Eur Psychiatry 2021; 64(Suppl 1): S63–4.
21. *Ahmed I, Ishtiaq S.* Reliability and validity: Importance in Medical Research. J Pak Med Assoc 2021; 71(10): 2401–6.
22. *Li F, Li D, Yu J, Jia Y, Jiang Y, Chen T,* et al. Barthel Index as a Predictor of Mortality in Patients with Acute Coronary Syndrome: Better Activities of Daily Living, Better Prognosis. Clin Interv Aging 2020; 15: 1951–61.
23. *Takabayashi K, Kitaguchi S, Iwatsu K,* Morikami Y, Ichinohe T, Yamamoto T, et al. A decline in activities of daily living due to acute heart failure is an independent risk factor of hospitalization for heart failure and mortality. J Cardiol 2019; 73(6): 522–9.
24. *Landis JR, Koch GG.* The measurement of observer agreement for categorical data. Biometrics 1977; 33(1): 159–74.
25. *Golden L, Rogers JG.* Heart Failure Drug Therapy: New Treatments, New Guidelines. Tex Heart Inst J 2022; 49(5): e227920.
26. *Spoto A, Nucci M, Prunetti E, Vicovaro M.* Improving content validity evaluation of assessment instruments through formal content validity analysis. Psychol Methods 2025; 30(2): 203–22.

Received on May 28, 2025
Revised on August 18, 2025
Accepted on September 24, 2025
Online First November 2025

Appendix 1.

The A-test – tabular presentation of the list used by physiotherapists [16,17].

A-test	Patient name					
	Diagnosis:					
	Day of rehabilitation					
1.	Bed mobility (from supine to side-lying)					
2.	From supine to sitting on the bed edge					
3.	Getting out of bed					
4.	Standing					
5.	Going back to bed					
6.	Walking with aids					
7.	Use of the toilet					
8.	Sitting on the chair					
9.	Walking up and down the stairs					
10.	Walking endurance					
	Summary					

Score: 0 – activity is not achieved; 1 – needs full assistance of a physiotherapist; 2 – requires adherence by a physiotherapist; 3 – activity performed with verbal suggestions of therapists; 4 – completely independent but insecure; 5 – fully independent and secure.

Score for walking endurance: 0 – unable to walk; 1 – walks up to 5 m (room-level walk); 2 – walks up to 15 m; 3 – walks up to 50 m; 4 – walks up to 100 m; 5 – walks over 100 m.