Physiotherapy Research International

WILEY

RESEARCH ARTICLE

Psychometric Properties of a Turkish Version of the Assessment of Physiotherapy Practice Tool

Furkan Cakir[1,2] | Ilksan Demirbuken[3] | Ender Ersin Avci[3] | Hulya Sisli[1] | Mine Gulden Polat[3] | Yavuz Yakut[4]

[1]Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Istanbul Bilgi University, Istanbul, Turkey | [2]Department of Physiotherapy and Rehabilitation, Institute of Health Sciences, Marmara University, Istanbul, Turkey | [3]Department of Physiotherapy Rehabilitation Education and Informatics, Faculty of Health Sciences, Marmara University, Istanbul, Turkey | [4]Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Hasan Kalyoncu University, Gaziantep, Turkey

Correspondence: Furkan Cakir (fztfurkancakir@gmail.com)

Received: 3 April 2025 | **Revised:** 14 August 2025 | **Accepted:** 23 September 2025

Funding: The authors received no specific funding for this work.

Keywords: clinical competence | educational measurement | internal consistency | intra-rater reliability | physical therapy | professional competence

ABSTRACT

Background and Purpose: Assessment of physiotherapy undergraduate students in clinical placement is academically important and holds practical value. This study aimed to translate the Assessment of Physiotherapy Practice (APP) tool, which is widely used for this purpose, into Turkish and to verify the factor validity and examine the reliability of this version using a cross-sectional design on senior physiotherapy students.

Methods: The APP and performance indicators were translated into Turkish in accordance with the recommended protocol. Exploratory and confirmatory factor analyses were conducted using scale data from 100 students. For reliability analyses, 10 clinical supervisors from a Turkish university assessed 63 students using the Turkish version of the Assessment of Physiotherapy Practice (APP-TR) tool. Supervisors performed the APP-TR assessment for each student at week 3 and at the end of the 6-week clinical placement, yielding a score for analysis.

Results: In both assessments, the scale demonstrated high levels of internal consistency (Cronbach's $\alpha = 0.961$ for the first assessment, 0.959 for final assessment). Two factors were identified by exploratory factor analysis explaining 65.84% of the total variance and a two-factor model was confirmed to fit by confirmatory factor analysis. Test-retest reliability was assessed by ICC and was high for all subheadings and total score. For all items, the close agreement was at least 98.41% and the exact agreement was at least 88.89% in the percentile analysis between the two assessments.

Discussion: The results of this study suggest that the APP-TR is a reliable and valid tool for evaluating final year physiotherapy students in a clinical placement in Turkey.

1 | Introduction

Physiotherapy education programs, like many other health disciplines (Chigavazira et al. 2018), contain clinical placements. Within these programs, workplace-based learning and assessment is recognised as an essential component. World Physiotherapy stipulates that practice education must account for approximately one third of the overall content of physiotherapy academic programs emphasizing its importance in physiotherapy education. Clinical education requires students to demonstrate competence in a spectrum of skills (e.g., communication, analysis, planning, etc.) and supervisors should include a variety of

approaches to student assessment (e.g., formative and summative) (World Physiotherapy 2021).

Considering the Physiotherapy entry level education requirements, the assessment of the student's clinical placement in various dimensions is both a quality indicator for the university and provides feedback to the student (Bithell 2007). In this area, valid and reliable assessment tools that assess psychometric properties appear to be necessary, and there is no standard measurement tool for evaluating the clinical practice performance of physiotherapy undergraduate students in Turkey.

Several clinical performance assessment tools have been utilized across different countries to appraise the proficiency of physiotherapy undergraduate students in their practices (Mori et al. 2015; Jones and Sheppard 2012; Reubenson et al. 2020). Among these clinical assessment tools, the Australian origin Assessment of Physiotherapy Practice (APP), which is widely used one, has been reported to have a high degree of criterion validity and reliability (O'Connor et al. 2018). The APP is a 20-item instrument covering Professional Behaviour (Item 1–4), Communication (Item 5–6), Assessment (Item 7–9), Analysis & Planning (Item 10–13), Intervention (Item 14–18), Evidence-based Practice (Item 19), and Risk Management (Item 20). There is an ''Examples of Performance Indicators'' section where all criteria are assessed. Students are expected to demonstrate these in their clinical practice. Each of the 20 criteria items has the following response options: '0 = Rarely demonstrates performance indicators', '1 = Demonstrates few performance indicators to an adequate standard', '2 = Demonstrates most performance indicators to an adequate standard', '3 = Demonstrates most performance indicators to a good standard', '4 = Demonstrates most performance indicators to an excellent standard'. For a parameter that cannot be assessed, the 'Not assessed' option is selected. The total score ranged from 0 to 80. A higher score indicates better performance. At the end of the form, there is a Global Rating Scale in which the student's overall clinical performance is selected as 'Not Adequate', 'Adequate', 'Good' and 'Excellent'. This scale is used to determine whether the student needs additional practice and is not included in the scoring (Dalton et al. 2011). This rubric-like structure made the tool stand out in terms of applicability and objectivity.

The cross-cultural applicability of the APP has been established through translation for use in China and Israel, and research has indicated the instrument's favourable psychometric properties across various nations (Hu et al. 2020; Anderson et al. 2014; Schwartz and Jacob 2019). Therefore, since there is no standard clinical performance assessment tool in our country, it is aimed to examine the adaptation, validity and reliability of the Turkish version of the APP (APP-TR).

2 | Methods

2.1 | Participants

The study was conducted on fourth-year physiotherapy students undertaking their clinical placements in orthopaedic physiotherapy during the 2020-2021 academic year at Marmara University, Turkey. Clinical supervisors were qualified physiotherapists with at least 3 years of experience affiliated with the institution's physiotherapy unit.

2.2 | Translation Procedure

After obtaining the necessary permissions, the APP and performance indicators were translated into Turkish in accordance with the recommended protocol (Guillemin et al. 1993). The protocol included the following steps: Forward translation was independently performed by two Turkish translators who were bilingual. These versions were discussed in a panel by two experts. Subsequently, two native English-speaking translators made independent back-translations. In another panel, compatibility with the original version was discussed. One of the panelists looked at professional semantic review, and the other looked at linguistic equivalence. The final version was created with the latest control and revisions. Instead of the direct translation of the word client in the phrase 'patient/client' in the scale, the equivalent of 'advisee' was linguistically preferred. No other changes were made in the panel where the final version was formed, and it was accepted as it is (See Supporting Information S1 for the final version of the APP-TR and performance indicators.)

2.3 | Data Collection and Analysis

This study was approved by Marmara University Health Sciences Institute's Ethics Committee (Protocol no: 12.2020-102). All participants gave written informed consent before data collection began. The research protocol was explained to the supervisors. Before the placements, students were briefed on the APP-TR tool and evaluation process. Prior to the study, all supervisors participated in face-to-face training on scoring assessment criteria and performance indicators. Student-supervisor pairings were randomized, and relevant reminders were made before the assessments. A total of 100 students were assessed by 10 supervisors, and the data from these assessments were used to conduct exploratory and confirmatory factor analyses. Additionally, another group of 63 students was evaluated by the same 10 supervisors during the third week and at the end of their 6-week clinical placements. The data from these evaluations were used to perform test-retest reliability analyses. Test and retest evaluations were conducted by the same supervisor who initially evaluated each student. The APP-TR tool scores were not taken into consideration while giving the clinical practice grades of the students. Assessment scales were taken by one of the researchers after the process was completed, and the identification information was removed before being entered into the spreadsheet for analysis.

2.4 | Statistical Analysis

Planned data analysis included: descriptive statistics, exploratory and confirmatory factor analysis, test-retest values, internal consistency analysis for total score and 7 subheadings, intraclass correlation coefficient (ICC) and confidence intervals for examining consistency between tests, calculation of Spearman's r for item total correlations, and percentile agreement graph between

tests. Analyses were performed using the SPSS (21.0 version) statistical software program (SPSS Inc. Chicago, USA), except confirmatory factor analysis (CFA) The CFA was carried out using SPSS Amos, version 21.0.

2.5 | Scale Characteristics

Scale median and min-max values, means and standard deviations for subheadings and total score and item-total correlations were calculated.

Internal consistency for each item, total score, and subheading scores of the Turkish scale was evaluated with Cronbach's alpha coefficient: $\alpha \geq 0.90$ was considered excellent; 0.80–0.90 good; 0.70–0.80 acceptable; 0.60–0.70 doubtful; and 0.50–0.60 poor (Torres-Narváez et al. 2018). To assess consistency between tests, the intraclass correlation coefficient (ICC) and corresponding confidence intervals were used. Based on the $ICC_{2,1}$ analysis, values between 0.60 and 0.80 were interpreted as indicating good reliability, while values exceeding 0.80 were deemed indicative of excellent reliability (Torres-Narváez et al. 2018). Additionally, the standard error of measurement (SEM) and minimal detectable change (MDC) were used for reliability analysis. The Spearman rank correlation with 95% confidence interval between each item in the first assessment and the total score was calculated. The cutoff values for rho were as follows: < 0.20 as poor correlation, 0.21 to ≤ 0.40 as fair, 0.41 to ≤ 0.60 as moderate, 0.61 to ≤ 0.80 as good, and 0.81 to ≤ 1.0 as very good (Shrout and Fleiss 1979). The percentage of agreement between the initial and final assessments is shown for each item in the APP-TR.

2.6 | Factorial Structure of the APP-TR

In the validation sample, exploratory factor analysis (EFA) was conducted using varimax rotation, an eigenvalue greater than one, and the scree test to determine the number of factors. The Kaiser-Meyer-Olkin (KMO) coefficient and Bartlett's test of sphericity were applied to evaluate the suitability of the dataset for factor analysis. A KMO value exceeding 0.60 and a statistically significant Bartlett's test result were deemed adequate for proceeding with factor analysis. In EFA, factor loadings of 0.45 or higher were considered a robust criterion for determining the alignment of items with their respective factors (Fitzgerald et al. 2007).

2.7 | Validation of Factorial Structure

In the same sample, confirmatory factor analysis (CFA) was performed. The overall model fit was assessed using the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI). Fit indices such as CFI > 0.95, TLI > 0.90, and RMSEA < 0.08 were regarded as indicative of a good model fit for continuous data (Schreiber et al. 2006). To verify the factorial structure, data from 100 assessed students were used, considering a 1:5 item-to-respondent ratio (Hair et al. 2018). 63 students enrolled in the entry-level physiotherapy program from a university in Istanbul were assessed by 10 supervisors using the APP-TR at the end of the 3rd and 6th weeks of the 6-week full-time clinical placement block scheduled across one university semester. For reliability analysis, a sample size was created similar to the original study (Dalton et al. 2012).

3 | Results

3.1 | Demographic Data

The sample for factor analysis comprised 100 students (65 females and 35 males) aged 21–24 years, along with 10 clinical supervisors (6 females and 4 males) from the physiotherapy unit. The reliability sample included 63 intern students (40 females and 23 males) within the same age range of 21–24 years. Both the

TABLE 1 | Test and retest results of APP-TR items and subheadings ($n = 63$).

	1st assessment (3rd week)		2nd assessment (6th week)	
Item number	**Median**	**Min–Max**	**Median**	**Min–Max**
Item 01	4	(1–4)	4	(1–4)
Item 02	4	(1–4)	4	(2–4)
Item 03	4	(2–4)	4	(2–4)
Item 04	4	(2–4)	4	(2–4)
Item 05	4	(2–4)	4	(2–4)
Item 06	4	(1–4)	4	(2–4)
Item 07	4	(1–4)	4	(2–4)
Item 08	3	(1–4)	3	(2–4)
Item 09	3	(1–4)	3	(2–4)
Item 10	3	(1–4)	3	(1–4)
Item 11	3	(1–4)	3	(1–4)
Item 12	3	(1–4)	3	(1–4)
Item 13	3	(1–4)	3	(1–4)
Item 14	3	(1–4)	3	(1–4)
Item 15	3	(1–4)	3	(1–4)
Item 16	3	(1–4)	3	(1–4)
Item 17	3	(1–4)	3	(1–4)
Item 18	3	(0–4)	3	(1–4)
Item 19	3	(1–4)	3	(2–4)
Item 20	3	(2–4)	3	(2–4)
Subheadings	**Mean**	**SD**	**Mean**	**SD**
Professional behaviour	14.03	2.16	14.16	2.03
Communication	6.70	1.44	6.86	1.28
Assessment	9.73	2.07	10.06	1.76
Analysis & planning	12.32	2.78	12.56	2.75
Intervention	15.40	3.50	15.63	3.29
Evidence-based practice	3.13	0.72	3.14	0.75
Risk management	3.05	0.68	3.08	0.65
Total	64.35	11.66	65.49	10.75

Abbreviations: Min–Max, Minimum–Maximum; SD, Standard Deviation.

supervisors and all students under their supervision participated in the assessment process. The clinical experience of the supervisors ranged between 3 and 20 years and their experience as a supervisor ranged between 3 and 16 years. The Turkish version of the APP was used for the first time by all supervisors.

3.2 | APP-TR Item and Subheading Scores

The APP item scores of 63 intern students were presented by giving their median and min-max values. The median value was 4 for the first seven items and 3 for the remaining items.

The first/final assessment scores were 14.03/14.16 for Professional Behaviour, 6.70/6.86 for Communication, 9.73/10.06 for Assessment, 12.32/12.56 for Analysis & Planning, 15.40/15.63 for Intervention, 3.13/3.14 for Evidence-based Practice and 3.05/3.08 for Risk Management. The mean total score was found to be 64.35 in the first assessment and 65.49 in the final assessment (Table 1).

3.3 | Construct Validity

Principal component analysis and the Varimax rotation method (for 20 items) were used in the EFA. The KMO = 0.939 and Bartlett's test of sphericity ($p < 0.001$) reached statistical significance, supporting the factor ability of the correlation matrix. In exploratory factor analysis, we found two factors with eigenvalues above 1. As a result of the exploratory factor analysis, it was determined that APP-TR has a two-factor structure, where the first factor explains 56.34% of the total variance (Item 7–20) and the second factor explains 9.503% of the total variance (Item 1–6). In cumulative total, all factors explain 65.84% of the total variance (Table 2).

Then, the two-factor model was tested by Confirmatory Factor Analysis (CFA). The CFA model is shown in Figure 1. In the confirmatory factor model, the CMIN/DF was 1.57, CFI was 0.963, TLI was 0.959 and RMSEA was 0.055 (90% confidence interval 0.032–0.074), indicating that the model fit the data well. The results of EFA and CFA showed that the structural validity of APP-TR was good.

3.4 | Reliability

The Cronbach's alpha value of the first test for APP-TR subheadings ranged from 0.832 to 0.897 and for the total score it was 0.961 (Table 3). In the final assessment, the Cronbach's alpha values for the APP-TR subheadings ranged from 0.808 to 0.912, and the alpha for the total score was 0.959 (Table 3).

TABLE 2 | Item loadings for principal component factor analysis and original scale dimension ($n = 100$).

	Factor loading[a]			
Item no	**Factor 1**	**Factor 2**	**Communality**	**Original scale dimension**
Item 01	0.382	**0.720**	0.710	Professional skills
Item 02	0.392	**0.775**	0.758	Professional skills
Item 03	0.323	**0.769**	0.704	Professional skills
Item 04	0.481	**0.803**	0.922	Professional skills
Item 05	0.114	**0.765**	0.417	Professional skills
Item 06	0.121	**0.719**	0.483	Professional skills
Item 07	**0.777**	0.215	0.611	Clinical skills
Item 08	**0.785**	0.243	0.435	Clinical skills
Item 09	**0.724**	0.298	0.456	Clinical skills
Item 10	**0.771**	0.169	0.557	Clinical skills
Item 11	**0.801**	0.216	0.533	Clinical skills
Item 12	**0.800**	0.188	0.550	Clinical skills
Item 13	**0.795**	0.263	0.678	Clinical skills
Item 14	**0.757**	0.251	0.590	Clinical skills
Item 15	**0.729**	0.289	0.573	Clinical skills
Item 16	**0.757**	0.330	0.663	Clinical skills
Item 17	**0.782**	0.311	0.686	Clinical skills
Item 18	**0.721**	0.308	0.606	Clinical skills
Item 19	**0.716**	0.349	0.455	Clinical skills
Item 20	**0.668**	0.296	0.504	Clinical skills
Percentage of variance	%56,34	%9,50		
Cumulative of variances	%65,84			

Note: Extraction method: Principal component analysis. Form items assigned to relevant factors with a factor loading above 0.45 are written in bold.
[a]Rotation converges after three iterations.

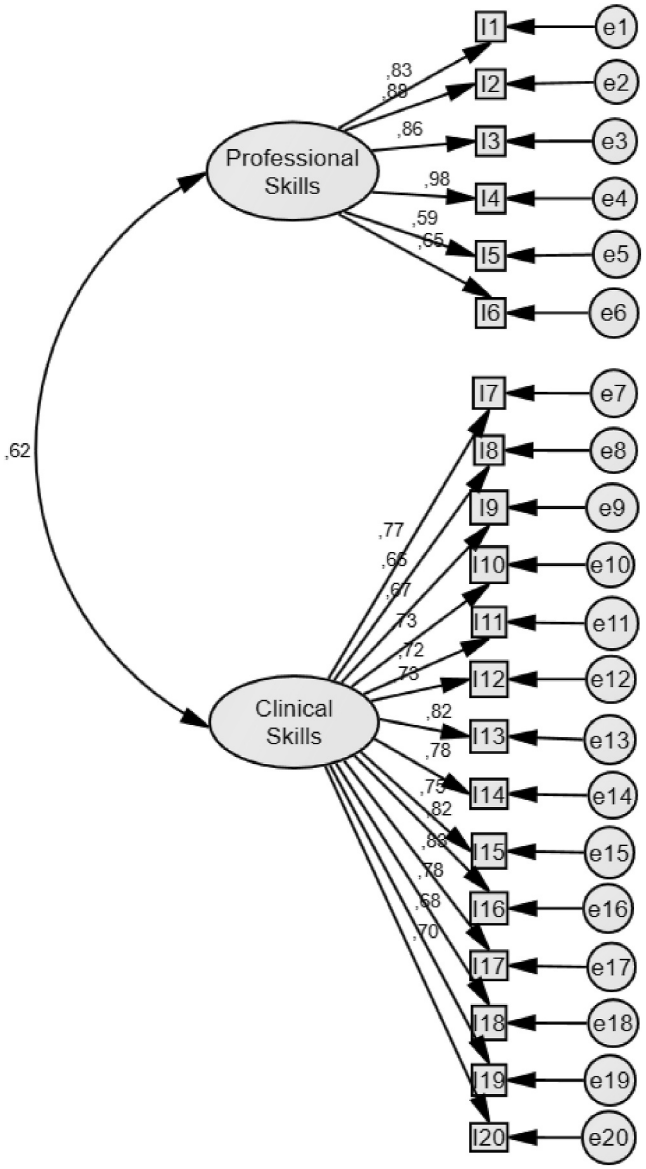

FIGURE 1 | A confirmatory factor model of the Turkish version of assessment of physiotherapy practice.

The $ICC_{2,1}$ of each item was found between 0.893 and 0.997 range, at the 95% confidence interval (Table 3). SEM and MDC results are presented in Table 4.

3.5 | Item-Total Correlations

In the item-total correlation analysis for 20 items in the scale in the first assessment, it was observed that the Spearman r value ranged between 0.659 and 0.845 ($p < 0.001$). In the final assessment at the 6th week, it was observed that the r value ranged between 0.641 and 0.857 ($p < 0.001$) (Table 5).

3.6 | Agreement Between Two Assessments

The present study evaluated the stability of the first and final assessment outputs of each criterion in the scale, and it was observed that close agreement ranged from 98.41% to 100%, and complete agreement ranged from 88.89% to 98.41% (Figure 2).

TABLE 3 | Cronbach's alpha for each subheading of the APP-TR ($n = 63$).

Clinical assessment categories	Cronbach's alpha	
	1st assessment (3rd week)	2nd assessment (6th week)
Professional behaviour (Items 1–4) Range: 0–16	0.874	0.889
Communication (Items 5–6) Range: 0–8	0.832	0.826
Assessment (Items 7–9) Range: 0–12	0.847	0.808
Analysis & planning (Items 10–13) Range: 0–16	0.897	0.912
Intervention (Items 14–18) Range: 0–20	0.894	0.889
Total APP-TR	0.961	0.959

4 | Discussion

The findings of the current study showed that the APP-TR version showed high internal consistency and excellent test-retest reliability. Factor analysis revealed two factors, similar to the original structure (Reubenson et al. 2020). Only items 5 and 6 (communication subheading items) were loaded on the professional skills factor and not on the clinical skill factor as expected. The tendency of the loadings in the Communication subheading (Items 5–6) to the Professional Skills factor may be due to the high correlation between communication skills and professional skills in the context of clinical training. This may be explained by the integration of communication skills with the perception of professional competence in the clinical culture in Turkey. The two-factor structure was retained in its current form and validated through confirmatory factor analysis, demonstrating a good level of fit based on the fit indices. The original scale was examined using Rasch analysis in previous studies (Dalton et al. 2011). In the present study, Rasch analysis was not conducted; instead, structural validity was assessed through factor analysis, and future research could use Rasch analysis to evaluate model fit and item-level measurement equivalence.

The results of this study are consistent with previous validation studies of the APP in various languages and contexts. A recent study conducted by Çelik et al. (2024) adapted the APP into Turkish and examined its psychometric and edumetric properties in physiotherapy education. Their findings demonstrated

TABLE 4 | Intraclass Correlation Coefficient (ICC) for items and each APP-TR subheading (n = 63).

Item number	$ICC_{2,1}$ (95% CI)	SEM	MDC
Item 01	0.981 (0.968–0.988)	0.03	0.08
Item 02	0.963 (0.939–0.978)	0.06	0.16
Item 03	0.986 (0.977–0.991)	0.01	0.02
Item 04	0.989 (0.982–0.993)	0.01	0.02
Item 05	0.951 (0.919–0.970)	0.07	0.19
Item 06	0.972 (0.955–0.983)	0.07	0.19
Item 07	0.935 (0.893–0.961)	0.11	0.30
Item 08	0.950 (0.918–0.970)	0.11	0.30
Item 09	0.954 (0.924–0.972)	0.11	0.30
Item 10	0.986 (0.976–0.991)	0.03	0.08
Item 11	0.976 (0.961–0.986)	0.04	0.11
Item 12	0.973 (0.956–0.984)	0.07	0.19
Item 13	0.972 (0.954–0.983)	0.07	0.19
Item 14	0.992 (0.987–0.995)	0.01	0.02
Item 15	0.989 (0.981–0.993)	0.03	0.08
Item 16	0.971 (0.952–0.982)	0.07	0.19
Item 17	0.970 (0.951–0.982)	0.06	0.16
Item 18	0.988 (0.980–0.993)	0.04	0.11
Item 19	0.950 (0.917–0.969)	0.07	0.19
Item 20	0.982 (0.971–0.989)	0.03	0.08
Subheadings	**$ICC_{2,1}$ (95% CI)**	**SEM**	**MDC**
Professional behaviour (Items 1–4),	0.992 (0.986–0.995)	0.12	0.33
Communication (Items 5–6)	0.977 (0.962–0.986)	0.15	0.41
Assessment (Items 7–9)	0.968 (0.948–0.981)	0.33	0.91
Analysis & planning (Items 10–13)	0.994 (0.990–0.996)	0.23	0.63
Intervention (Items 14–18)	0.995 (0.991–0.997)	0.23	0.63
Evidence-based practice (Item 19)	0.950 (0.917–0.969)	0.07	0.19
Risk management (Item 20)	0.982 (0.971–0.989)	0.03	0.08
Total APP	0.995 (0.992–0.997)	1.20	3.32

Abbreviations: CI, Confidence Interval; MDC, Minimal Detectable Change; SEM, Standard Error of Measurement.

TABLE 5 | Item total correlations for APP-TR (n = 63).

Item total correlations	1st assessment (3rd week)		2nd assessment (6th week)	
Item number	**rho**	***p***	**rho**	***p***
Item 01	0.696	< 0.001	0.686	< 0.001
Item 02	0.774	< 0.001	0.740	< 0.001
Item 03	0.659	< 0.001	0.661	< 0.001
Item 04	0.758	< 0.001	0.756	< 0.001
Item 05	0.675	< 0.001	0.641	< 0.001
Item 06	0.741	< 0.001	0.752	< 0.001
Item 07	0.732	< 0.001	0.711	< 0.001
Item 08	0.845	< 0.001	0.785	< 0.001
Item 09	0.765	< 0.001	0.721	< 0.001
Item 10	0.763	< 0.001	0.740	< 0.001
Item 11	0.772	< 0.001	0.810	< 0.001
Item 12	0.841	< 0.001	0.823	< 0.001
Item 13	0.822	< 0.001	0.857	< 0.001
Item 14	0.774	< 0.001	0.776	< 0.001
Item 15	0.723	< 0.001	0.742	< 0.001
Item 16	0.816	< 0.001	0.769	< 0.001
Item 17	0.791	< 0.001	0.782	< 0.001
Item 18	0.759	< 0.001	0.786	< 0.001
Item 19	0.704	< 0.001	0.727	< 0.001
Item 20	0.691	< 0.001	0.685	< 0.001

Note: rho = Spearman Correlation Coefficient.

high internal consistency (Cronbach's $\alpha > 0.90$) and strong inter-rater reliability, further supporting the applicability of the APP within the Turkish clinical education context (Çelik et al. 2024). Although methodological differences exist between the two studies, both contribute to the growing body of evidence on the reliability and validity of the Turkish version of the APP. The quality of applied education and its assessment process is an issue that needs to be emphasized for clinical educators. In Turkey, studies examining standardized clinical performance assessment tools in physiotherapy education are still very limited. Apart from the recent work of Çelik et al. (2024), no other study has comprehensively evaluated such tools within our context. Our findings therefore provide additional evidence on the applicability and measurement properties of the APP-TR and complement the existing literature.

The descriptive statistics of the study demonstrate that the median value of the initial seven criteria is 4, whereas that of the subsequent items is 3. The high median value in the Professional Behaviour and Communication headings revealed the scores where the skills of the students in Turkey are high. It was observed that the scores were relatively low in Analysis & Planning, Intervention, Evidence-based Practice and Risk Management. It may be beneficial to support students more in these subjects and to focus on the aspects that can be developed individually in clinical training.

The test-retest reliability was as high as 0.99 in the ICC for the total APP-TR score. The results of the original scales' test-retest reliability study revealed an $ICC_{2,1}$ of 0.92 (95% CI, 0.84–0.96) with a standard error of measurement of 3.2 points on an 80-point scale and a minimal detectable change (at 90% confidence) of 7.86 points (Murphy et al. 2014).

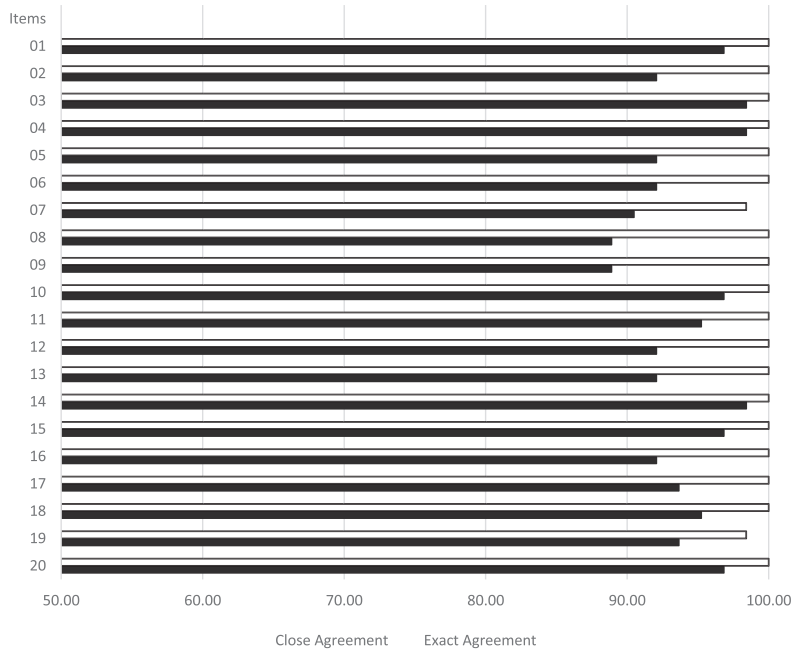

FIGURE 2 | Percentage of agreement for test and retest.

The high retest correlations shown in the original version and the APP-TR provide evidence that educators using the APP are consistent in rating students' relative abilities. This is important for monitoring improvement in performance relative to peers. The study conducted on the original scale indicated that the margin of error (95% CI) was acceptable using a scale width of 0–80, and the test-retest correlations were high, similar to the findings of our study (Dalton et al. 2012). This error provides a high level of accuracy in ranking student performance, as evidenced by the test/retest correlation of 0.99.

Internal consistency was examined for the first 5 subheadings of the scale and it was observed that it was at a high level for both test and retest. It has been stated that the Cronbach's alpha value of the Hebrew version of the scale is between 0.65 and 0.93 (Schwartz and Jacob 2019). In our study, the lowest Cronbach's alpha value was found to be 0.935, and the highest was 0.995.

Item total correlations were found to be above 0.60 in the first and last evaluations for all items and it was accepted as strong. In the literature review, close findings of an assessment tool with a similar structure were also shown (Torres-Narváez et al. 2018; Dalton et al. 2012). Based on our observations during the study, neither the students being evaluated nor the supervisors reported any difficulties with the use of the APP-TR or the assessment process.

The limitations of this study include the absence of comparable instruments for external validity testing, as it is, to the best of our knowledge, the first standardised tool validated in Turkish to assess the clinical practice competence of physiotherapy and rehabilitation undergraduate students. Furthermore, the findings are based on data from a single university with a relatively small sample size, which may limit the generalizability of the results. Further studies conducted at different universities will contribute to the broad validity and reliability of the APP-TR.

5 | Implications for Physiotherapy Practice

According to the study findings, the APP-TR demonstrated satisfactory factor validity, good internal consistency, and excellent test-retest reliability in assessing the skills and competencies of clinical placement students, making it suitable for this purpose. This scale may guide the development of new assessment tools by helping to determine and structure student performance criteria during clinical practice. Future research should examine its applicability across all fields of physiotherapy and its use in broader populations.

Acknowledgements

The authors have nothing to report.

Ethics Statement

The Marmara University Health Sciences Institute's Ethics Committee approved the study and provided approval for consent waiver (Protocol no:12.2020-102).

Consent

All participants signed an informed consent form before registration.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Data available upon request.

Permission to Reproduce Material From Other Sources

The authors have permission to reproduce the material contained in this study.

References

Anderson, C., M. Cosgrove, D. Lees, et al. 2014. "What Clinical Instructors Want: Perspectives on a New Assessment Tool for Students in the Clinical Environment." *Physiotherapy Canada* 66, no. 3: 322–328. https://doi.org/10.3138/ptc.2013-27.

Bithell, C. 2007. "Entry-Level Physiotherapy Education in the United Kingdom: Governance and Curriculum." *Physical Therapy Reviews* 12, no. 2: 145–155. https://doi.org/10.1179/108331907X175041.

Çelik, H. İ., D. Ozturk, M. Sari, N. Bulut, and N. Bek. 2024. "Psychometric and Edumetric Properties of the Turkish Version of the Assessment of Physiotherapy Practice." *BMC Medical Education* 24, no. 1: 1162. https://doi.org/10.1186/s12909-024-06180-w.

Chigavazira, J., R. Fernandez, M. Mackay, and S. Lapkin. 2018. "Adaptation and Validation of the Clinical Supervision Self-Assessment Tool Among Registered Nurses." *Nurse Education Today* 70: 28–33. https://doi.org/10.1016/j.nedt.2018.08.008.

Dalton, M., M. Davidson, and J. L. Keating. 2011. "The Assessment of Physiotherapy Practice (APP) Is a Valid Measure of Professional Competence of Physiotherapy Students: A Cross-Sectional Study With Rasch Analysis." *Journal of Physiotherapy* 57, no. 4: 239–246. https://doi.org/10.1016/S1836-9553(11)70054-6.

Dalton, M., M. Davidson, and J. L. Keating. 2012. "The Assessment of Physiotherapy Practice (APP) Is a Reliable Measure of Professional Competence of Physiotherapy Students: A Reliability Study." *Journal of Physiotherapy* 58, no. 1: 49–56. https://doi.org/10.1016/S1836-9553(12)70072-3.

Fitzgerald, L. M., A. Delitto, and J. J. Irrgang. 2007. "Validation of the Clinical Internship Evaluation Tool." *Physical Therapy* 87, no. 7: 844–860. https://doi.org/10.2522/ptj.20060054.

Guillemin, F., C. Bombardier, and D. Beaton. 1993. "Cross-Cultural Adaptation of Health-Related Quality of Life Measures: Literature Review and Proposed Guidelines." *Journal of Clinical Epidemiology* 46, no. 12: 1417–1432. https://doi.org/10.1016/0895-4356(93)90142-N.

Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. 2018. *Multivariate Data Analysis*. 8th ed. Cengage Learning.

Hu, J., A. Y. Jones, X. Zhou, et al. 2020. "Acceptance of the 'Assessment of Physiotherapy Practice (Chinese)' as a Standardised Evaluation of Professional Competency in Chinese Physiotherapy Students: An Observational Study." *BMC Medical Education* 20, no. 1: 1–7. https://doi.org/10.1186/s12909-020-02026-3.

Jones, A., and L. Sheppard. 2012. "Developing a Measurement Tool for Assessing Physiotherapy Students' Self-Efficacy: A Pilot Study." *Assessment & Evaluation in Higher Education* 37, no. 3: 369–377. https://doi.org/10.1080/02602938.2010.534765.

Mori, B., D. Brooks, K. Norman, J. Herold, and D. Beaton. 2015. "Development of the Canadian Physiotherapy Assessment of Clinical Performance: A New Tool to Assess Physiotherapy Students' Performance in Clinical Education." *Physiotherapy Canada* 67, no. 3: 281–289. https://doi.org/10.3138/ptc.2014-29E.

Murphy, S., M. Dalton, and D. Dawes. 2014. "Assessing Physical Therapy Students' Performance During Clinical Practice." *Physiotherapy Canada* 66, no. 2: 169–176. https://doi.org/10.3138/ptc.2013-26.

O'Connor, A., O. McGarr, P. Cantillon, A. McCurtin, and A. Clifford. 2018. "Clinical Performance Assessment Tools in Physiotherapy Practice Education: A Systematic Review." *Physiotherapy* 104, no. 1: 46–53. https://doi.org/10.1016/j.physio.2017.01.005.

Reubenson, A., L. Ng, and D. F. Gucciardi. 2020. "The Assessment of Physiotherapy Practice Tool Provides Informative Assessments of Clinical and Professional Dimensions of Student Performance in Undergraduate Placements: A Longitudinal Validity and Reliability Study." *Journal of Physiotherapy* 66, no. 2: 113–119. https://doi.org/10.1016/j.jphys.2020.03.009.

Schreiber, J. B., A. Nora, F. K. Stage, E. A. Barlow, and J. King. 2006. "Reporting Structural Equation Modeling and Confirmatory Factor Analysis Results: A Review." *Journal of Educational Research* 99, no. 6: 323–338. https://doi.org/10.3200/JOER.99.6.323-338.

Schwartz, D., and T. Jacob. 2019. "Establishing the Reliability of a Tool for Assessing Israeli Physical Therapy Students' Clinical Performance." *Journal, Physical Therapy Education* 33, no. 3: 243–248. https://doi.org/10.1097/JTE.0000000000000093.

Shrout, P. E., and J. L. Fleiss. 1979. "Intraclass Correlations: Uses in Assessing Rater Reliability." *Psychological Bulletin* 86, no. 2: 420–428. https://doi.org/10.1037/0033-2909.86.2.420.

Torres-Narváez, M. R., O. C. Vargas-Pinilla, and E. I. Rodríguez-Grande. 2018. "Validity and Reproducibility of a Tool for Assessing Clinical Competencies in Physical Therapy Students." *BMC Medical Education* 18, no. 1: 1–8. https://doi.org/10.1186/s12909-018-1377-x.

World Physiotherapy. 2021. Physiotherapist Education Framework [Internet]. https://world.physio/sites/default/files/2021-07/Physiotherapist-education-framework-FINAL.pdf.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Supporting Information S1: pri70114-sup-0001-suppl-data.doc.