Journal of Physiotherapy 66 (2020) 113–119

Journal of
PHYSIOTHERAPY

journal homepage: www.elsevier.com/locate/jphys

Research

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

Alan Reubenson [a], Leo Ng [a], Daniel F Gucciardi [a,b]

[a] School of Physiotherapy and Exercise Science, Curtin University, Perth, Australia; [b] Physical Activity and Well-being Research Group, Curtin University, Perth, Australia

KEY WORDS

Exploratory structural equation modelling
Physical therapy
Professional competence
Psychometric
Clinical education

ABSTRACT

Questions: Do one or two factors best represent clinical performance scores obtained via the Assessment of Physiotherapy Practice (APP) and what is the nature of their characterisation? To what extent are the same number of factors and their interpretation, and item scaling captured equally over time and across contexts (eg, clinical subdisciplines) for assessments of clinical performance via the APP? **Design**: Archival and longitudinal study of undergraduate students' clinical performances for each of four final-year clinical placements. **Participants**: A total of 561 undergraduate physiotherapy students from one Australian university who were enrolled to complete their final-year clinical placements between 2014 and 2017. **Outcome measures**: Clinical educators' assessments of student performance across seven key domains of clinical practice: professional behaviour, communication, assessment, analysis and planning, intervention, evidence-based practice and risk management. **Results**: Factor analyses supported the superiority of a two-factor representation of the APP, including dimensions characterised by professional and clinical domains, when compared with a unidimensional structure of an overarching 'clinical performance' factor. It was also found that the two-factor representation and item scaling was consistent across four clinical placements covering typical areas of physiotherapy practice. In other words, the same constructs are being assessed equally well across context and time. **Conclusions**: The APP is the nationally adopted assessment tool that is used to evaluate clinical competence to practise as a physiotherapist in Australia and New Zealand. These findings provide new evidence for an updated scoring protocol in which clinical factors are distinguished from professional competencies. **[Reubenson A, Ng L, Gucciardi DF (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* 66:113–119]**

© 2020 Australian Physiotherapy Association. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Physiotherapy training, like many other healthcare disciplines, contains clinical components alongside traditional academic requirements encountered in other schooling such as non-health-related courses.[1] The clinical components demand competence in areas that span cognitive skills (eg, understanding of concepts), physical skills (eg, manual therapy techniques) and interpersonal skills (eg, communication). Students are also required to make higher-level clinical reasoning decisions, integrating theory and skills into practice in a safe manner. Clinical performance is judged by qualified professionals who take into consideration these many factors and a 'gut-feel' approach is often used to evaluate student performance in this setting.[2] The availability of a tool that provides reliable and valid evaluations of clinical performance is therefore critical to the standardisation of assessments over time and across clinical subdisciplines.

The Assessment of Physiotherapy Practice (APP) is the nationally adopted assessment tool used to evaluate clinical competence to practise as a physiotherapist in Australia and New Zealand. (For a review of available tools, see the review by O'Connor et al.[3]) Underpinned by the Australian Standards of Physiotherapy Practice,[4,5] the APP contains 20 items to determine entry-level competence across seven key domains of clinical practice. Each item is scored using a 5-point scale to create a total score out of 80, which is considered alongside a global rating score (not adequate, adequate, good or excellent). Typically, clinical educators provide a single evaluation of students on these items from multiple observations of their performance across a clinical placement lasting 4 to 6 weeks. Students' entry-level clinical competence is assessed across several placements in different subdisciplines of practice (eg, musculoskeletal and cardiopulmonary) and settings (eg, acute, community and rehabilitation). To be deemed competent, students must achieve a total score > 50% and meet the

https://doi.org/10.1016/j.jphys.2020.03.009
1836-9553/© 2020 Australian Physiotherapy Association. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

minimal entry-level competence standard on the global rating score.

The APP was originally developed to provide a unidimensional assessment of entry-level clinical competence.[6] Psychometric evaluations of the APP by the tool creators provided support for a unidimensional structure[7] and a high-level of inter-rater reliability (ICC = 0.92, 95% CI 0.82 to 0.96) between paired assessors who evaluated 30 students' performances over a 5-week clinical placement.[8] This foundational work led to the APP being adopted as the primary assessment tool of entry-level clinical competence within tertiary training programs in Australia and New Zealand, including utilisation within simulation-based education.[9] However, despite the widespread adoption of the APP, little research on the psychometric properties of this tool has been published since its original development. There is potential for further rigorous testing of the APP because the psychometric properties of a tool (eg, factor structure and internal reliability) may differ according to context and time.[10] Measurement invariance – evidence that the same construct(s) is being captured equally well over time and context – is of particular relevance for entry-level clinical competence because students are evaluated across several clinical placements in different subdisciplines over 12 to 18 months.

Broadly speaking, psychometric evaluations can focus on the substance (eg, extent to which items adequately capture professional practice), structure (eg, number of unique factors captured in the APP) and/or external nature (eg, degree to which entry-level clinical performance predicts employability) of a tool.[11] Support for the substance of clinical performance as assessed via the APP was provided in the original development work (ie, content validity evidence)[6] and in the widespread use of the tool among tertiary institutions in Australia and New Zealand (ie, face validity). With regard to scale structure, Dalton et al[7] provided initial validity evidence for a unidimensional representation of the 20 items. In contrast to this evidence, tertiary institutions (Clinical Education Managers of Australia and New Zealand 2020, personal communication) and researchers[12,13] have subsequently used alternative scoring protocols for the APP, where there are specific scores for professional/generic skills (items 1 to 6) and clinical skills (items 7 to 20). The absence of independent tests of the structural properties of the APP beyond the original work, when combined with varied applications in practice and research, underscores the need for a rigorous examination of the factorial structure of this tool. Such data are important for two reasons. First, clarification of the structural properties of the APP can inform educational practice (eg, evidence-based scoring protocol) and policy (eg, standardisation of the use of a national tool) in positive ways. Second, the availability of a robust operationalisation has the potential to maximise confidence in and the accumulation of evidence regarding key determinants and outcomes of clinical performance.

In summary, the aim of this study was to examine the structural properties of the APP among a cohort of undergraduate physiotherapy students undertaking final-year clinical placements.

Therefore, the research questions for this archival and longitudinal study were:

1. Do one or two factors best represent clinical performance scores obtained via the Assessment of Physiotherapy Practice and what is the nature of their characterisation?
2. To what extent are the same number of factors and their interpretation, and item scaling captured equally over time and across contexts (eg, clinical subdisciplines) for assessments of clinical performance via the Assessment of Physiotherapy Practice?

Method

Design

In this archival and longitudinal study, clinical educators assessed each student's clinical performance for each of four final-year clinical placements. (All other placements during the undergraduate degree are not assessed using the APP.) All of the examined placements are 5-week units where students engage in applying knowledge and skills in real-world settings across core areas, including musculoskeletal, neurology, cardiopulmonary and lifespan (eg, paediatrics and gerontology). Clinical supervisors assess students based on their performance across the entire placement against entry-level competencies aligned with the Australia and New Zealand physiotherapy standards using the APP tool.

Participants

Participants in the study were undergraduate physiotherapy students from one Australian university who were enrolled to complete their final-year clinical placements between 2014 and 2017. Students were excluded if they had completed only one final-year placement before exiting the course.

Outcome measures

The APP is a 20-item tool that covers seven key domains of clinical practice: professional behaviour; communication; assessment; analysis and planning; intervention; evidence-based practice; and risk management. Each item is scored using a 5-point scale (0 = demonstrates the performance criterion infrequently/rarely to 4 = an excellent standard). Interested readers are referred to Dalton et al for a detailed overview of its development and original psychometric validation.[6]

Data analysis

Key details of the primary analyses are described below. All M*plus* syntax scripts and output are available on the Open Science Framework (OSF; http://bit.ly/APPstudyJoP).

Factorial validity evidence

In phase one, the factor structure of the APP was tested using confirmatory factor analysis (CFA) and exploratory structural equation modelling (ESEM)[14] with robust maximum likelihood estimation in M*plus* 8.3.[15] Briefly, CFA enables hypothesised factor structures to be tested in a way that presumes that items are perfect indicators of latent constructs, whereas ESEM allows imprecision in measurement indicators via the estimation of cross-loadings on non-target factors. A review and comparison of these analytical approaches is available elsewhere.[16,17] A unidimensional structure was examined as per the original protocol[6] and compared against a two-factor model, as used in recent work[12,13] across each measurement wave (ie, time point) and across each cognate group (ie, clinical subdiscipline) separately. A visual depiction of this difference and the models tested is provided in Figure 1. Target rotations were used in all ESEM analyses to model cross-loadings on non-target factors to be as close to 0 as possible.[14,18]

In phase two, the measurement invariance of the best fitting model from phase one was tested via a commonly accepted three-step approach[19,20] to determine if the (i) number of latent factors and items per factor (configural), (ii) strength of the factor loadings (metric) and (iii) strength of the factor loadings and item intercepts (scalar) are equivalent over time and clinical subdiscipline (ie, cardiopulmonary, musculoskeletal, neurological and alternative or miscellaneous placement). Model-data fit was assessed using a multifaceted approach including the Chi-squared goodness-of-fit index, comparative fit index (CFI), Tucker-Lewis index (TLI) and root mean square error of approximation (RMSEA). Values of CFI and TLI $\geq$ 0.90 and RMSEA $\leq$ 0.08 were considered to indicate acceptable fit.[21] Reliability of latent factors was estimated using omega coefficient (ω).[22] For comparisons among nested models in measurement invariance tests, a scaled Chi-squared difference test[23] was employed alongside changes in model-data fit indices, such that the more complex model was favoured when ΔCFI $<$ 0.01 and ΔRMSEA $<$ 0.015.[24]

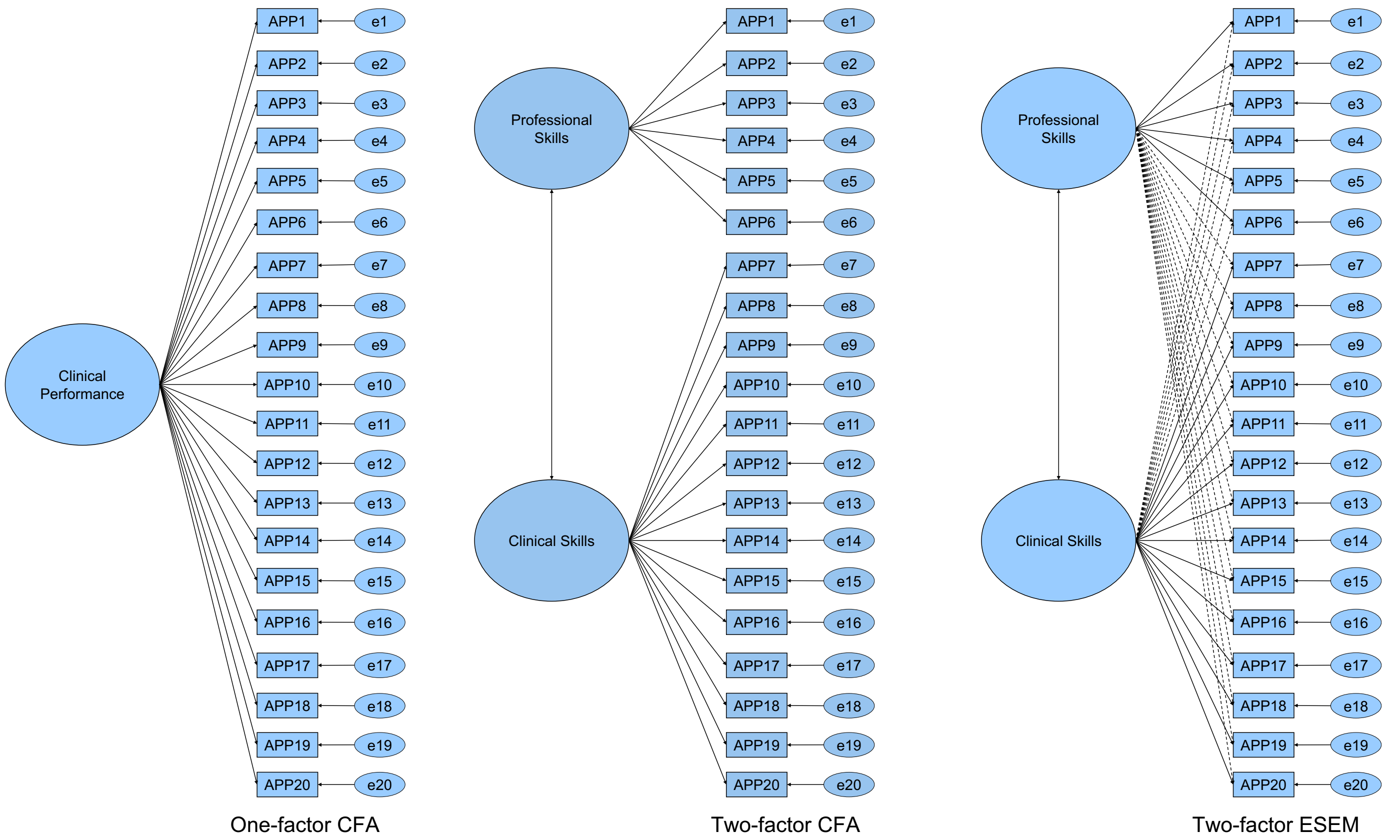

Figure 1. Visual depiction of one-factor and two-factor models for confirmatory factor analyis (CFA) and exploratory structural equation modelling (ESEM).
Elipses (circles) represent latent variables, boxes represent observed variables, single-headed arrows represent a directional effect of a latent variable on an observed variable, double-headed arrows represent correlation among latent factors, solid lines represent target factor loading and dashed lines represent non-target factor loading.
APP = Assessment of Physiotherapy Practice item, e = residual variance.

Results

Flow of participants

In total, 564 students were eligible for inclusion in this study, three of whom were deemed ineligible. Of the 561 retained participants, five students were assessed on two placements, 14 were assessed on three placements and 542 were assessed on four placements. Baseline demographic information is provided in Table 1.

Factorial and measurement invariance validity

Item-level descriptive statistics for each assessment point indicated that the data were approximately normally distributed (see output files located on the OSF: http://bit.ly/APPstudyJoP). Model-data fit indices for CFA and ESEM supported the superiority of the two-factor structure, including latent factors characterised by 'clinical' and 'professional' dimensions, when compared with a unidimensional representation of APP scores by measurement wave (see Table 2) and cognate area (see Table 3). Latent factor reliability estimates were excellent for all models ($\omega > 0.80$).[25] The standardised factor loadings were excellent for the two-factor CFA model ($\Lambda > 0.60$; see output files located on the OSF). Standardised loadings largely supported well-defined factors in the two-factor ESEM structure with regard to target loadings ($\Lambda > 0.45$); however, three items (APP5, APP6 and APP20) loaded poorly on their target factor ($\Lambda < 0.25$) and meaningfully on the non-target factor ($\Lambda > 0.45$). Subsequently, a revised model was tested in which the professional factor is characterised by items 1 to 4 because the item descriptions and performance indicators provide substantive grounds to consider them as indicators of professional skills. This representation demonstrated better model-data fit relative to the two-factor model in which professional skills are characterised by items 1 to 6 (see Tables 2 and 3). Standardised loadings supported well-defined factors in this revised two-factor ESEM structure with regard to the target factor and non-target factor (see Tables 4 and 5). This two-factor CFA model was preferable as the best fitting solution for its parsimony.[26] Tests of longitudinal measurement invariance supported configural, metric and scalar invariance for the two-factor CFA and ESEM models over time (see Table 6). These findings indicate that the number of latent factors, meaning ascribed to the latent factors, and scaling of item scores are the same over four clinical placements. In contrast, tests of cognate area invariance supported the equivalency of the number of latent factors and the meaning associated with them across the four clinical subdisciplines but

Table 1
Demographic information of participants.

Characteristic	Participants (n = 561)
Age (*y*), mean (SD)	23 (4)
Gender, n (%)	
female	378 (67)
male	183 (33)
Citizenship, n (%)	
Australia	490 (87)
Singapore	25 (5)
Malaysia	19 (3)
other	27 (5)
Main language, n (%)	
English	513 (91)
Cantonese	8 (1)
Mandarin	6 (1)
other	34 (6)
High school, n (%)	
Australia	499 (89)
international	62 (11)

Table 2
Summary of fit indices for measurement models by measurement wave (ie, first to fourth clinical placements).

	χ^2	df	p	AIC	BIC	BIC[adj]	RMSEA (90% CI)	CFI	TLI
CFA: one-factor (wave 1)	940.52	170	< 0.001	16503.97	16762.46	16571.99	0.091 (0.085 to 0.097)	0.899	0.887
CFA: one-factor (wave 2)	826.12	170	< 0.001	16670.61	16930.07	16739.60	0.083 (0.078 to 0.089)	0.914	0.904
CFA: one-factor (wave 3)	958.72	170	< 0.001	16297.75	16557.00	16366.52	0.091 (0.086 to 0.097)	0.898	0.885
CFA: one-factor (wave 4)	1126.48	170	< 0.001	16379.79	16639.14	16448.67	0.101 (0.095 to 0.106)	0.882	0.868
CFA: two-factor[a] (wave 1)	744.79	169	< 0.001	16277.06	16539.86	16346.22	0.079 (0.073 to 0.085)	0.924	0.915
CFA: two-factor[a] (wave 2)	661.14	169	< 0.001	16478.22	16742.01	16548.36	0.072 (0.066 to 0.078)	0.936	0.928
CFA: two-factor[a] (wave 3)	699.58	169	< 0.001	16001.82	16265.39	16071.75	0.075 (0.069 to 0.081)	0.931	0.922
CFA: two-factor[a] (wave 4)	827.71	169	< 0.001	16041.49	16305.16	16111.52	0.084 (0.078 to 0.089)	0.919	0.909
CFA: two-factor[b] (wave 1)	624.49	169	< 0.001	16141.29	16404.09	16210.45	0.070 (0.064 to 0.076)	0.940	0.933
CFA: two-factor[b] (wave 2)	571.81	169	< 0.001	16376.54	16640.32	16446.68	0.065 (0.060 to 0.071)	0.947	0.941
CFA: two-factor[b] (wave 3)	613.85	169	< 0.001	15905.28	16168.84	15975.20	0.069 (0.063 to 0.075)	0.942	0.935
CFA: two-factor[b] (wave 4)	731.46	169	< 0.001	15936.65	16200.33	16006.69	0.077 (0.072 to 0.083)	0.931	0.922
ESEM: two-factor[a] (wave 1)	511.32	151	< 0.001	16048.19	16388.53	16137.76	0.066 (0.060 to 0.072)	0.953	0.940
ESEM: two-factor[a] (wave 2)	374.56	151	< 0.001	16189.14	16530.77	16279.98	0.052 (0.045 to 0.058)	0.971	0.963
ESEM: two-factor[a] (wave 3)	421.47	151	< 0.001	15725.99	16067.33	15816.54	0.057 (0.050 to 0.063)	0.965	0.956
ESEM: two-factor[a] (wave 4)	526.47	151	< 0.001	15752.16	16093.65	15842.86	0.067 (0.061 to 0.073)	0.954	0.942
ESEM: two-factor[b] (wave 1)	511.32	151	< 0.001	16048.19	16388.53	16137.76	0.066 (0.060 to 0.072)	0.953	0.940
ESEM: two-factor[b] (wave 2)	374.56	151	< 0.001	16189.14	16530.77	16279.98	0.052 (0.045 to 0.058)	0.971	0.963
ESEM: two-factor[b] (wave 3)	421.47	151	< 0.001	15725.99	16067.33	15816.54	0.057 (0.050 to 0.063)	0.965	0.956
ESEM: two-factor[b] (wave 4)	526.47	151	< 0.001	15752.16	16093.65	15842.86	0.067 (0.061 to 0.073)	0.954	0.942

AIC = Akaike information criterion, BIC = Bayesian information criterion, BIC[adj] = sample size adjusted BIC, CFA = confirmatory factor analysis, CFI = comparative fit index, df = degrees of freedom, ESEM = exploratory structural equation model with target rotations, RMSEA = root mean square error of approximation, TLI = Tucker-Lewis index, χ^2 = Chi-square.
[a] f1 (items 1 to 6) and f2 (items 7 to 20).
[b] f1 (items 1 to 4) and f2 (items 5 to 20).

not the scaling of item scores. Subsequent exploratory analyses supported partial scalar invariance, where items 12 (sets realistic short-term and long-term client-centred goals), 18 (undertakes discharge planning) and 20 (identifies adverse events/near misses and minimises risk associated with assessment and interventions) are allowed to differ across cognate areas (see http://bit.ly/APPstudyJoP). Considered in conjunction with item means, partial scalar invariance indicated that assessments for musculoskeletal placements on these three indicators are higher than the average, whereas cardiopulmonary placements are lower.

Discussion

This study is the first longitudinal investigation of the psychometric properties of the nationally adopted assessment tool that is used to evaluate clinical competency to practise as a physiotherapist in Australia and New Zealand. It found that clinical supervisors' assessments of undergraduate physiotherapy students' performances using the APP are best characterised by two factors representing clinical and professional skills. It also found that the interpretation of the two-factor representation and item scaling was largely consistent across four clinical placements covering typical areas of physiotherapy practice.

Knowledge of the factor structure of assessment tools is essential information for their appropriate use in practice, particularly for high-stake assessments of students' clinical performance. A recent systematic review found that no clinical performance assessment tool in physiotherapy education meets the highest standard of evidence, when considered against several key criteria of high-quality assessment methods.[3] The evidence is insufficient, which is largely due to limited ongoing tests of reliability and validity evidence of test scores obtained using such tools. Most pertinent to the current study, there has been only one subsequent

Table 3
Summary of fit indices for measurement models by clinical subdiscipline.

	χ^2	df	p	AIC	BIC	BIC[adj]	RMSEA (90% CI)	CFI	TLI
CFA: one-factor (A)	761.54	170	< 0.001	16228.91	16487.39	16296.93	0.080 (0.074 to 0.085)	0.928	0.919
CFA: one-factor (C)	788.90	170	< 0.001	16764.57	17024.67	16834.20	0.080 (0.075 to 0.086)	0.928	0.919
CFA: one-factor (M)	1279.32	170	< 0.001	15165.45	15424.48	15234.01	0.109 (0.103 to 0.114)	0.826	0.805
CFA: one-factor (N)	1159.79	170	< 0.001	16774.52	17033.44	16842.97	0.103 (0.097 to 0.108)	0.875	0.861
CFA: two-factor[a] (A)	607.49	169	< 0.001	16055.30	16318.10	16124.46	0.069 (0.063 to 0.075)	0.947	0.940
CFA: two-factor[a] (C)	700.98	169	< 0.001	16664.57	16929.01	16735.37	0.075 (0.069 to 0.080)	0.938	0.930
CFA: two-factor[a] (M)	964.23	169	< 0.001	14779.59	15042.94	14849.30	0.092 (0.087 to 0.098)	0.875	0.860
CFA: two-factor[a] (N)	735.50	169	< 0.001	16297.17	16560.41	16366.77	0.078 (0.072 to 0.084)	0.929	0.920
CFA: two-factor[b] (A)	568.40	169	< 0.001	16010.85	16273.64	16080.00	0.066 (0.060 to 0.072)	0.951	0.945
CFA: two-factor[b] (C)	642.88	169	< 0.001	16598.76	16863.20	16669.56	0.071 (0.065 to 0.076)	0.945	0.938
CFA: two-factor[b] (M)	798.14	169	< 0.001	14599.33	14862.68	14669.04	0.082 (0.076 to 0.088)	0.901	0.889
CFA: two-factor[b] (N)	641.89	169	< 0.001	16194.28	16457.52	16263.88	0.071 (0.065 to 0.077)	0.940	0.933
ESEM: two-factor[a] (A)	386.86	151	< 0.001	15842.76	16183.10	15932.32	0.053 (0.047 to 0.060)	0.971	0.964
ESEM: two-factor[a] (C)	541.48	151	< 0.001	16516.01	16858.48	16607.69	0.068 (0.062 to 0.074)	0.954	0.943
ESEM: two-factor[a] (M)	573.14	151	< 0.001	14390.07	14731.12	14480.34	0.071 (0.065 to 0.077)	0.934	0.917
ESEM: two-factor[a] (N)	441.41	151	< 0.001	16006.04	16346.95	16096.17	0.059 (0.053 to 0.065)	0.963	0.954
ESEM: two-factor[b] (A)	386.86	151	< 0.001	15842.76	16183.10	15932.32	0.053 (0.047 to 0.060)	0.971	0.964
ESEM: two-factor[b] (C)	541.48	151	< 0.001	16516.01	16858.48	16607.69	0.068 (0.062 to 0.074)	0.954	0.943
ESEM: two-factor[b] (M)	573.14	151	< 0.001	14390.07	14731.12	14480.34	0.071 (0.065 to 0.077)	0.934	0.917
ESEM: two-factor[b] (N)	441.41	151	< 0.001	16006.04	16346.95	16096.17	0.059 (0.053 to 0.065)	0.963	0.954

A = alternative (ie, miscellaneous), AIC = Akaike information criterion, BIC = Bayesian information criterion, BIC[adj] = sample size adjusted BIC, C = cardiopulmonary, CFA = confirmatory factor analysis, CFI = comparative fit index, df = degrees of freedom, ESEM = exploratory structural equation model with target rotations, M = musculoskeletal, N = neurological, RMSEA = root mean square error of approximation, TLI = Tucker-Lewis index, χ^2 = Chi-square.
[a] f1 (items 1 to 6) and f2 (items 7 to 20).
[b] f1 (items 1 to 4) and f2 (items 5 to 20).

Table 4
Standardised factor loadings of two-factor measurement models for confirmatory factor analysis and exploratory structural equation models by measurement wave (factor 1 = items 1 to 4 and factor 2 = items 5 to 20).

	Wave 1				Wave 2				Wave 3				Wave 4			
	C1	C2	E1	E2	C1	C2	E1	E2	C1	C2	E1	E2	C1	C2	E1	E2
APP1	0.85[a]		0.94[a]	−0.09[a]	0.84[a]		0.88[a]	−0.01	0.85[a]		0.89[a]	−0.02	0.86[a]		0.84[a]	0.05
APP2	0.76[a]		0.53[a]	0.26[a]	0.75[a]		0.44[a]	0.36[a]	0.78[a]		0.55[a]	0.27[a]	0.82[a]		0.65[a]	0.20[a]
APP3	0.89[a]		0.96[a]	−0.06	0.82[a]		0.84[a]	0.01	0.86[a]		0.90[a]	−0.02	0.87[a]		0.88[a]	0.01
APP4	0.74[a]		0.47[a]	0.30[a]	0.78[a]		0.54[a]	0.27[a]	0.78[a]		0.52[a]	0.29[a]	0.77[a]		0.54[a]	0.28[a]
APP5		0.74[a]	0.11	0.66[a]		0.75[a]	0.18[a]	0.62[a]		0.71[a]	0.25[a]	0.52[a]		0.75[a]	0.27[a]	0.54[a]
APP6		0.68[a]	0.11	0.69[a]		0.62[a]	0.09	0.55[a]		0.66[a]	0.14[a]	0.55[a]		0.66[a]	0.08	0.60[a]
APP7		0.79[a]	0.04	0.76[a]		0.77[a]	0.09	0.71[a]		0.77[a]	0.04	0.84[a]		0.79[a]	0.06	0.75[a]
APP8		0.79[a]	0.05	0.75[a]		0.83[a]	−0.05	0.87[a]		0.83[a]	−0.07	0.89[a]		0.79[a]	−0.10[a]	0.86[a]
APP9		0.80[a]	0.13[a]	0.69[a]		0.84[a]	−0.02	0.85[a]		0.82[a]	−0.01	0.83[a]		0.81[a]	−0.10[a]	0.89[a]
APP10		0.83[a]	−0.05	0.87[a]		0.85[a]	−0.08[a]	0.91[a]		0.82[a]	−0.13[a]	0.93[a]		0.82[a]	−0.12[a]	0.91[a]
APP11		0.80[a]	−0.05	0.84[a]		0.83[a]	−0.12[a]	0.92[a]		0.83[a]	−0.16[a]	0.96[a]		0.83[a]	−0.07	0.89[a]
APP12		0.75[a]	−0.09	0.82[a]		0.76[a]	0.08	0.70[a]		0.71[a]	0.05	0.67[a]		0.73[a]	0.01	0.73[a]
APP13		0.83[a]	−0.06	0.89[a]		0.84[a]	−0.03	0.87[a]		0.82[a]	−0.02	0.84[a]		0.81[a]	−0.06	0.86[a]
APP14		0.83[a]	0.04	0.79[a]		0.83[a]	−0.05	0.86[a]		0.85[a]	−0.01	0.85[a]		0.82[a]	−0.06	0.87[a]
APP15		0.75[a]	0.00	0.75[a]		0.78[a]	−0.02	0.79[a]		0.75[a]	0.08	0.69[a]		0.78[a]	−0.04	0.80[a]
APP16		0.82[a]	−0.15[a]	0.95[a]		0.81[a]	−0.13[a]	0.92[a]		0.83[a]	−0.04	0.86[a]		0.83[a]	−0.03	0.86[a]
APP17		0.82[a]	−0.08	0.89[a]		0.80[a]	−0.07	0.85[a]		0.81[a]	−0.10	0.88[a]		0.81[a]	−0.02	0.83[a]
APP18		0.74[a]	0.01	0.73[a]		0.73[a]	0.17[a]	0.61[a]		0.76[a]	0.12[a]	0.67[a]		0.74[a]	0.09	0.67[a]
APP19		0.69[a]	0.12	0.60[a]		0.72[a]	0.21[a]	0.55[a]		0.70[a]	0.07	0.64[a]		0.67[a]	0.28[a]	0.45[a]
APP20		0.75[a]	0.26[a]	0.54[a]		0.68[a]	0.48[a]	0.31[a]		0.68[a]	0.44[a]	0.35[a]		0.71[a]	0.48[a]	0.35[a]
ω	0.88	0.96	0.87	0.96	0.87	0.96	0.84	0.96	0.89	0.96	0.87	0.96	0.90	0.96	0.87	0.96
ψ	0.80		0.77		0.81		0.73		0.79		0.73		0.79		0.73	

APP = Assessment of Physiotherapy Practice item, C = confirmatory factor analysis, E = exploratory structural equation model with target rotations, ψ = latent variable correlation, ω = omega estimate of internal reliability, blue shading = intended factor loading on target dimension.
[a] $p < 0.05$.

or independent examination of the psychometric properties of the APP[9] since the original development and validation work.[7,8] That study[9] addressed that gap in the evidence by conducting a rigorous assessment of the APP among a large sample of physiotherapy students across two universities. Contrary to the original development work of the APP[6,7] and the only independent validation study,[9] we found that a two-factor structure provides the best representation of supervisors' assessments of students' clinical performances in the field. The two-factor structure was largely robust across four final-year clinical placements, which provides users with confidence that the same constructs are being assessed equally well across time and context. Nevertheless, if researchers or clinical educators wish to compare latent factors between cognate areas, the findings indicate that efforts need to be made to account for minor differences in item scaling for three items (items 12, 18 and 20).

The current evidence supports APP scoring protocols used in recent work,[12,13] yet raises the possibility of an alternative composition of the two factors representing clinical and professional factors. Our data support the superiority of a professional

Table 5
Standardised factor loadings of two-factor measurement models for confirmatory factor analysis and exploratory structural equation models by cognate area (factor 1 = items 1 to 4 and factor 2 = items 5 to 20).

	Alternative				Cardiopulmonary				Musculoskeletal				Neurological			
	C1	C2	E1	E2	C1	C2	E1	E2	C1	C2	E1	E2	C1	C2	E1	E2
APP1	0.82[a]		0.91[a]	−0.05	0.81[a]		0.91[a]	−0.07	0.89[a]		0.94[a]	−0.04	0.88[a]		0.86[a]	0.05
APP2	0.77[a]		0.47[a]	0.32[a]	0.77[a]		0.40[a]	0.38[a]	0.71[a]		0.47[a]	0.35[a]	0.82[a]		0.70[a]	0.14[a]
APP3	0.82[a]		0.84[a]	0.01	0.85[a]		0.97[a]	−0.09	0.86[a]		0.90[a]	−0.05	0.89[a]		0.85[a]	0.06
APP4	0.78[a]		0.42[a]	0.39[a]	0.77[a]		0.38[a]	0.40[a]	0.75[a]		0.63[a]	0.16[a]	0.77[a]		0.60[a]	0.22[a]
APP5		0.77[a]	0.24[a]	0.58[a]		0.75[a]	0.14	0.64[a]		0.65[a]	0.12[a]	0.57[a]		0.76[a]	0.38[a]	0.54[a]
APP6		0.72[a]	0.13[a]	0.62[a]		0.63[a]	−0.05	0.67[a]		0.55[a]	0.09	0.49[a]		0.69[a]	0.22[a]	0.52[a]
APP7		0.81[a]	0.13[a]	0.70[a]		0.81[a]	0.10	0.72[a]		0.73[a]	−0.04	0.76[a]		0.79[a]	0.11[a]	0.71[a]
APP8		0.89[a]	0.01	0.80[a]		0.83[a]	−0.02	0.85[a]		0.77[a]	−0.10[a]	0.84[a]		0.81[a]	−0.06	0.86[a]
APP9		0.83[a]	0.10[a]	0.75[a]		0.84[a]	0.04	0.80[a]		0.77[a]	−0.08[a]	0.83[a]		0.83[a]	0.00	0.83[a]
APP10		0.85[a]	−0.06	0.89[a]		0.83[a]	−0.06	0.88[a]		0.79[a]	−0.08[a]	0.84[a]		0.86[a]	−0.13[a]	0.96[a]
APP11		0.83[a]	−0.02	0.84[a]		0.83[a]	−0.04	0.86[a]		0.79[a]	−0.14[a]	0.88[a]		0.86[a]	−0.13[a]	0.96[a]
APP12		0.78	−0.05	0.82[a]		0.73[a]	−0.08	0.80[a]		0.68[a]	0.05	0.65[a]		0.74[a]	−0.02	0.75[a]
APP13		0.85[a]	−0.10[a]	0.93[a]		0.85[a]	−0.05	0.89[a]		0.81[a]	−0.04	0.84[a]		0.79[a]	−0.02	0.81[a]
APP14		0.86[a]	−0.01	0.87[a]		0.85[a]	0.10[a]	0.76[a]		0.80[a]	−0.04	0.82[a]		0.82[a]	0.00	0.82[a]
APP15		0.78[a]	0.04	0.75[a]		0.78[a]	0.12	0.68[a]		0.72[a]	0.02	0.71[a]		0.77[a]	−0.05	0.80[a]
APP16		0.84[a]	−0.18[a]	0.98[a]		0.85[a]	−0.10[a]	0.94[a]		0.78[a]	0.10[a]	0.72[a]		0.83[a]	−0.11[a]	0.91[a]
APP17		0.84[a]	−0.11[a]	0.93[a]		0.80[a]	−0.04	0.84[a]		0.80[a]	−0.02	0.81[a]		0.80[a]	−0.10[a]	0.87[a]
APP18		0.79[a]	−0.03	0.82[a]		0.76[a]	0.04	0.72[a]		0.70[a]	0.22[a]	0.56[a]		0.69[a]	0.07	0.64[a]
APP19		0.70[a]	0.25[a]	0.50[a]		0.75[a]	0.18[a]	0.60[a]		0.67[a]	0.14[a]	0.58[a]		0.62[a]	0.10	0.54[a]
APP20		0.75[a]	0.37[a]	0.45[a]		0.79[a]	0.25[a]	0.57[a]		0.52[a]	0.41[a]	0.27[a]		0.68[a]	0.48[a]	0.33[a]
ω	0.87	0.97	0.82	0.96	0.88	0.96	0.83	0.96	0.88	0.95	0.86	0.94	0.90	0.96	0.88	0.96
ψ	0.85		0.77		0.88		0.83		0.64		0.60		0.77		0.71	

APP = Assessment of Physiotherapy Practice item, C = confirmatory factor analysis, E = exploratory structural equation model with target rotations, ψ = latent variable correlation, ω = omega estimate of internal reliability, blue shading = intended factor loading on target dimension.
[a] $p < 0.05$.

Table 6
Summary of fit indices for measurement invariance analyses.

	χ^2	df	p	RMSEA (90% CI)	CFI	TLI	ΔRMSEA	ΔCFI	ΔTLI
CFA – longitudinal									
configural invariance	2537.72	676	<0.001	0.070 (0.068 to 0.073)	0.940	0.933	–	–	–
metric invariance	2603.26	730	<0.001	0.068 (0.065 to 0.071)	0.940	0.937	0.002	0.000	0.004
scalar invariance	2697.48	784	<0.001	0.066 (0.064 to 0.069)	0.938	0.940	0.002	0.002	0.003
configural versus metric[a]	51.47	54	0.57	–	–	–	–	–	–
metric versus scalar[a]	84.97	54	0.005	–	–	–	–	–	–
CFA – cognate area									
configural invariance	2651.32	676	<0.001	0.073 (0.070 to 0.075)	0.936	0.929	–	–	–
metric invariance	2803.09	730	<0.001	0.072 (0.069 to 0.074)	0.933	0.931	0.001	0.003	0.002
scalar invariance	3434.11	784	<0.001	0.078 (0.075 to 0.081)	0.915	0.917	0.006	0.018	0.014
configural versus metric[a]	147.32	54	<0.001	–	–	–	–	–	–
metric versus scalar[a]	670.39	54	<0.001	–	–	–	–	–	–
ESEM – longitudinal									
configural invariance	1831.71	604	<0.001	0.061 (0.057 to 0.064)	0.960	0.950	–	–	–
metric invariance	1985.03	712	<0.001	0.057 (0.054 to 0.060)	0.959	0.956	0.004	0.001	0.006
scalar invariance	2073.58	766	<0.001	0.055 (0.053 to 0.058)	0.958	0.958	0.002	0.001	0.002
configural versus metric[a]	149.73	108	<0.001	–	–	–	–	–	–
metric versus scalar[a]	99.80	54	<0.001	–	–	–	–	–	–
ESEM – cognate area									
configural invariance	1943.50	604	<0.001	0.063 (0.060 to 0.066)	0.957	0.946	–	–	–
metric invariance	2213.03	712	<0.001	0.062 (0.059 to 0.065)	0.952	0.948	0.001	0.005	0.002
scalar invariance	2795.29	766	<0.001	0.069 (0.066 to 0.072)	0.935	0.935	0.007	0.017	0.013
configural versus metric[a]	264.47	108	<0.001	–	–	–	–	–	–
metric versus scalar[a]	491.31	54	<0.001	–	–	–	–	–	–

CFA = confirmatory factor analysis, CFI = comparative fit index, df = degrees of freedom, ESEM = exploratory structural equation model with target rotations, RMSEA = root mean square error of approximation, TLI = Tucker-Lewis index, χ^2 = Chi-square.
The two-factor model tested for invariance was f1 (items 1 to 4) and f2 (items 5 to 20).
[a] = chi square difference test using the Satorra and Bentler (2001) method (http://www.statmodel.com/chidiff.shtml).

dimension that is characterised by an understanding of client rights and consent; commitment to learning; ethical, legal and culturally sensitive practice; and teamwork. This representation differs from past scoring protocols that incorporate communication and clear and accurate documentation as part of the professional factor.[12,13] Communication as a concept is broad and complex (eg, interpersonal style and non/verbal), particularly within the context of therapist-client relationships, where both the substance and style are critical.[27] Exemplars of performance indicators for APP assessments provided to clinical supervisors incorporate the accuracy of what is said (eg, clear instructions) and how that information is delivered or received (eg, listens carefully). Ultimately, however, the majority of these exemplars rely on sound clinical knowledge and skills to maximise the quality of this communication. Asking appropriate questions to gather relevant information, for example, requires that students possess adequate knowledge and understanding of various clinical presentations to make informed decisions and apply evidence-based practice (ie, quality in equals quality out). Similarly, this consideration applies equally to medico-legal documentation as captured in item six. Collectively, therefore, our empirical data alongside substantive interpretations of item content provide an update to the APP scoring protocol for consideration in future research and practice.

Key strengths of this study include a large sample, longitudinal design, rigorous statistical analyses and availability of annotated syntax. Nevertheless, there are two key limitations that can inform future research, namely the need to test the robustness of these findings across different sites, geographical locations and supervisors' demographics (eg, professional experience); and examine other types of validity evidence (eg, predictive). Despite the national adoption of the APP as the primary competency-based assessment of clinical performance in physiotherapy practice education, implementation varies in educational practice. Thus, this study addresses an important gap in the literature by conducting a rigorous psychometric examination of the APP. These findings provide evidence to institutions and clinical educators that the APP is most appropriately scored and interpreted via two factors encompassing professional (items 1 to 4) and clinical (items 5 to 20) skills.

What was already known on this topic: The Assessment of Physiotherapy Practice (APP) is the nationally adopted assessment tool for the evaluation of clinical competence to practise as a physiotherapist in Australia and New Zealand. The APP contains 20 items that cover seven key domains of clinical practice.
What this study adds: Factor analyses supported the superiority of a two-factor representation of the APP, including dimensions characterised by professional (items 1 to 4) and clinical (items 5 to 20) domains. The two-factor representation and item scaling were consistent across four clinical placements covering typical areas of physiotherapy practice.

Ethics approval: Curtin University's Human Research Ethics Committee approved the study (HRE2017-0568) and provided approval for consent waiver.
Competing interests: Nil.
Source(s) of support: Nil.
Acknowledgements: Nil.
Provenance: Not invited. Peer reviewed.
Correspondence: Alan Reubenson, School of Physiotherapy and Exercise Science, Curtin University, Australia. Email: A.Reubenson@curtin.edu.au

References

1. Thieman TJ, Weddle ML, Moore MA. Predicting academic, clinical, and licensure examination performance in a professional (entry-level) Master's degree program in physical therapy. *J Phys Ther Educ*. 2003;17:32–37.
2. Jette DU, Bertoni A, Coots R, Johnson H, McLaughlin C, Weisbach C. Clinical instructors' perceptions of behaviors that comprise entry-level clinical performance in physical therapist students: a qualitative study. *Phys Ther*. 2007;87:833–843.
3. O'Connor A, McGarr O, Cantillon P, McCurtin A, Clifford A. Clinical performance assessment tools in physiotherapy practice education: a systematic review. *Physiotherapy*. 2018;104:46–53.
4. Australian Physiotherapy Council. Accreditation Standard Requirements. 2016. https://apc-11666.kxcdn.com/wp-content/uploads/2019/08/ACCREDITATION-STANDARD-V1.1-13112017.pdf. Accessed 28 January, 2020.

5. Physiotherapy Board of Australia & Physiotherapy Board of New Zealand. Physiotherapy practice thresholds in Australia and Aotearoa New Zealand. 2015. https://physiocouncil.com.au/media/1020/physiotherapy-board-physiotherapy-practice-thresholds-in-australia-and-aotearoa-new-zealand-6.pdf. Accessed 28 January, 2020.
6. Dalton M, Keating J, Davidson M. Development of the Assessment of Physiotherapy Practice (APP): A standardised and valid approach to assessment of clinical competence in physiotherapy. [Australian Learning and Teaching Council (ALTC) Final report PP6-28]. 2009. http://www.altc.edu.au/resource-development-clinical-assessment-griffith-2009. Accessed 28 January, 2020
7. Dalton M, Davidson M, Keating J. The Assessment of Physiotherapy Practice (APP) is a valid measure of professional competence of physiotherapy students: a cross-sectional study with Rasch analysis. *J Physiother*. 2011;57:239–246.
8. Dalton M, Davidson M, Keating JL. The Assessment of Physiotherapy Practice (APP) is a reliable measure of professional competence of physiotherapy students: a reliability study. *J Physiother*. 2012;58:49–56.
9. Judd BK, Scanlan JN, Alison JA, Waters D, Gordon CJ. The validity of a professional competence tool for physiotherapy students in simulation-based clinical education: a Rasch analysis. *BMC Med Educ*. 2016;16:196.
10. Flora DB, Flake JK. The purpose and practice of exploratory and confirmatory factor analysis in psychological research: decisions for scale development and validation. *Can J Behav Sci*. 2017;49:78–88.
11. Loevinger J. Objective tests as instruments of psychological theory. *Psychol Rep*. 1957;3:635–694.
12. Blackford J, McAllister L, Alison J. Simulated learning in the clinical education of novice physiotherapy students. *Int J Pract Based Learn Health Social Care*. 2015;3:77–93.
13. Wright A, Moss P, Dennis DM, Harrold M, Levy S, Furness AL, et al. The influence of a full-time, immersive simulation-based clinical placement on physiotherapy student confidence during the transition to clinical practice. *Adv Simul*. 2018;3:3.
14. Asparouhov T, Muthén B. Exploratory structural equation modeling. *Struct Equ Modeling*. 2009;16:397–438.
15. Muthén LK, Muthén BO. *Mplus User's Guide*. 8th ed. Los Angeles: Muthén & Muthén; 1998–2017.
16. Gucciardi DF, Zyphur MJ. Exploratory structural equation modelling and Bayesian estimation. In: Ntoumanis N, Myers ND, eds. *An Introduction to Intermediate and Advanced Statistical Analyses for Sport and Exercise Scientists*. Chichester: Wiley; 2016:172–194.
17. Marsh HW, Morin AJ, Parker PD, Kaur G. Exploratory structural equation modeling: an integration of the best features of exploratory and confirmatory factor analysis. *Annu Rev Clin Psychol*. 2014;10:85–110.
18. Browne MW. An overview of analytic rotation in exploratory factor analysis. *Multivariate Behav Res*. 2001;36:111–150.
19. Dimitrov DM. Testing for factorial invariance in the context of construct validation. *Meas Eval Couns Dev*. 2010;43:121–149.
20. Vandenberg RJ, Lance CE. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ Res Methods*. 2000;3:4–70.
21. Marsh HW, Hau K-T, Grayson D. Goodness of fit in structural equation models. In: *Contemporary psychometrics: A festschrift for Roderick P. McDonald*. New Jersey: Lawrence Erlbaum Associates; 2005:275–340.
22. McDonald RP. The theoretical foundations of principal factor analysis, canonical factor analysis and alpha factor analysis. *Br J Math Stat Psychol*. 1970;23:1–21.
23. Satorra A, Bentler PM. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*. 2001;66:507–514.
24. Marsh HW, Ludtke O, Muthen B, Asparouhov T, Morin AJ, Trautwein U, et al. A new look at the big five factor structure through exploratory structural equation modeling. *Psychol Assess*. 2010;22:471–491.
25. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. *Br J Psychol*. 2014;105:399–412.
26. Marsh HW, Nagengast B, Morin AJS. Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and la dolce vita effects. *Dev Psychol*. 2013;49:1194–1218.
27. Pinto RZ, Ferreira ML, Oliveira VC, Franco MR, Adams R, Maher CG, et al. Patient-centred communication is associated with positive therapeutic alliance: a systematic review. *J Physiother*. 2012;58:77–87.