Clinical Performance Assessment Tools in Physiotherapy Practice Education; A Systematic Review

O’Connor, A.1Anne.OConnor@ul.ie,
McGarr, O.2Oliver.McGarr@ul.ie,
Cantillon, P.3Peter.Cantillon@nuigalway.ie,
McCurtin, A.1Arlene.McCurtin@ul.ie,
Clifford, A.1Amanda.Clifford@ul.ie

1Department of Clinical Therapies, Health Sciences Building, University of Limerick, Limerick, Ireland.
2Department of Education and Professional Studies, University of Limerick, Limerick, Ireland
3Discipline of General Practice, Clinical Science Institute, National University of Ireland, Galway, Ireland.

*Corresponding authors at: Department of Clinical Therapies, Health Sciences Building, University of Limerick, Limerick, Ireland. Tel.: +353 61 233279/ +353 86 8523977, Fax: +353 61 234251
Abstract

Background: Clinical performance assessment tools (CPATs) used in physiotherapy practice education need to be psychometrically sound and appropriate for use in all clinical settings in order to provide an accurate reflection of a student’s readiness for clinical practice. Current evidence to support the use of existing assessment tools is inconsistent.

Objectives: To conduct a systematic review synthesising evidence relating to the psychometric and edumetric properties of CPATS used in physiotherapy practice education.

Data Sources: An electronic search of Web of Science, SCOPUS, Academic Search Complete, AMED, Biomedical Reference Collection, British Education Index, CINAHL plus, Education Full Text, ERIC, General Science Full Text, Google Scholar, MEDLINE, UK and Ireland Reference Centre databases was conducted identifying English language papers published in this subject area from 1985 to 2015.

Study selection: Twenty papers were identified representing fourteen assessment tools.

Data Extraction and Synthesis: Two reviewers evaluated selected papers using a validated framework (Swing et al., 2009).

Results: Evidence of psychometric testing was inconsistent and varied in quality. Reporting of edumetric properties was unpredictable in spite of its importance in busy clinical environments. No Class 1 recommendation was made for any of the CPATs, and no CPAT scored higher than Level C evidence.

Conclusions: Findings demonstrate poor reporting of psychometric and edumetric properties of CPATs reviewed. A more robust approach is required when designing CPATs. Collaborative endeavour within the physiotherapy profession and interprofessionally may be key to further developments in this area and may help strengthen the rigour of such assessment processes.

Keywords: clinical performance assessment, physical therapy, physiotherapy, student, assessment tool.
Contribution of Paper:

This systematic review identifies and synthesises the evidence relating to the psychometric and edumetric properties of clinical performance assessment tools used in physiotherapy practice education.

Findings highlight that psychometric and edumetric evidence of these assessment tools is reported inconsistently, and these properties require more systematic and rigorous testing procedures in the early stages of tool development.

Collaborative research effort inclusive of physiotherapy and other health professions may help provide a greater bank of knowledge in clinical performance assessment.

Introduction

The World Confederation of Physical Therapy (WCPT) stipulate that practice education must account for approximately one third of the overall content of physiotherapy academic programmes (1, 2) emphasising its importance in physiotherapy education. Physiotherapy students must “meet the competencies established by the physical therapist professional entry level education programme” (1) and must be provided with formative and summative feedback during each practice education module (2). This is achieved through an assessment process, where clinical performance is assessed based on observation by a supervising clinician, known as a practice educator.

Clinical performance assessment has long challenged education providers for reasons related to evidence supporting assessment methods and factors related to the subjective nature of observation-based assessment (3-9). No literature review to date has synthesised the evidence related to psychometric testing (validity and reliability) and edumetric properties (feasibility, usefulness and educational impact) of CPATs used in physiotherapy practice education.
Current evidence suggests that psychometric evidence for many of these is inconsistent (10-12) with little or no attention paid to their edumetric properties.

A recent systematic review in medical education also acknowledged poor reporting of edumetric properties in CPATs (13). This is despite their importance in determining a tool’s practicality and feasibility in the workplace. Lengthy or ambiguously worded assessment tools can frustrate busy clinicians which in turn can impact on rigorous completion of student assessments (14). Psychometric properties are more commonly reported although not always comprehensively (12, 13). Such properties include content validity which captures how accurately learning outcomes described in a CPAT measure various aspects of clinical performance. This is determined by matching selected assessment criteria with published guidelines required for physiotherapy entry level practice (15, 16). Criterion validity assesses the extent to which the measure is related to the outcomes. Evidence of construct validity demonstrates that a tool is sensitive enough to detect changes in student performance over time which confirms progression of student learning (16-18). Additional psychometric properties include inter-rater and intra-rater reliability which when present ensure consistency of grading across a variety of practice educators and practice education sites. Test-retest reliability is also necessary in CPATs to ensure consistent ranking of students on repeated assessment, particularly relevant in practice education when regular observation forms the basis for awarding final grades. Therefore, CPATs with less than acceptable psychometric and edumetric testing may cast doubt on their inherent ability to identify both the excelling student and the incompetent or unsafe student. This can result in an assessment process that is potentially unreliable and precarious with implications for educational programmes, client safety and professional bodies (3, 6).
Physiotherapy undergraduate students, like nursing students, are expected to work unsupervised from the time of graduation unlike medical students who must complete further postgraduate study, known as an internship, in order to practice independently. Only one qualitative evaluation framework has been developed to evaluate and synthesise evidence pertaining to the psychometric and edumetric properties of clinical performance assessment tools used in the clinical learning environment. This was developed by the Accreditation Council for Graduate Medical Education (ACGME) in the United States of America (19). This framework defined guidelines for grading psychometric and edumetric properties of assessment instruments as well as outlining a system for assigning an overall evidence grade for each tool. While developed for graduate medical students, it was considered appropriate for use in this study as it lends itself easily to use by other health professions (13) especially given the similarities in expectations of the physiotherapy graduate and the medical intern.

In the absence of other reviews of this kind, the need to evaluate and synthesise the evidence related to the edumetric and psychometric properties of CPATs used in physiotherapy education was deemed essential.

The specific aims of this systematic review were to:

1. Identify and synthesise available evidence pertaining to the psychometric and edumetric properties of clinical performance assessment tools used in physiotherapy practice education using the ACGME framework.
2. Discuss the findings within the broader context of health professional clinical performance assessment.
Methods

Search Strategy

A systematic review of the literature was undertaken using combinations of search terms (Table 1) to identify English language peer-reviewed papers published from January 1985 to December 2015 relating to CPATs used in physiotherapy. Prior to 1985 few clinical performance assessment tools had been reported in physiotherapy literature (20, 21).

Databases included Web of Science, Academic Search Complete, AMED, Biomedical Reference Collection, British Education Index, CINAHL plus, Education Full Text, ERIC, General Science Full Text, MEDLINE, UK and Ireland Reference Centre, Google Scholar and SCOPUS. Reference lists were examined by hand for further citations and checked for eligibility using the same inclusion and exclusion criteria.

INSERT TABLE 1 HERE

Inclusion Criteria

- Any peer-reviewed paper describing a CPAT used in physiotherapy practice education employing observation-based assessment methods, which included reference to psychometric or edumetric testing.
- Experimental and observational studies, randomised and non-randomised designs, and prospective or retrospective cohort studies.
- Full text papers written in the English language
- Publication date from January 1985 to December 2015.
Exclusion Criteria

- Any peer-reviewed paper describing assessment tools where standardised patients, simulated settings, Objective Structured Clinical Examinations or learning portfolios were used to assess clinical performance.
- Assessment tools exclusively from disciplines other than physiotherapy.
- Research studies where the full-text paper was not available.
- Assessment tools involving student self-assessment only.
- Audits based on the use of an assessment tool but without reference to psychometric testing or edumetric properties.
- The development of a tool but no evidence of testing or validation.
- Assessment tools used solely for physiotherapy postgraduate education where specialisation in an area of physiotherapy practice was the subject of assessment.

Study Selection & Data Extraction

An electronic search was completed by reviewer (W). Titles and abstracts of studies retrieved were independently screened to identify studies meeting the inclusion criteria. A system was used to label these studies. Those labelled ‘yes’ and ‘unsure’ were checked independently by reviewer (X) for decision regarding inclusion. If a decision was unclear from the abstract, the full-text article was obtained for assessment. Any disagreements were resolved through discussion between the reviewers (W and X). A further reviewer (Y) was available for disagreements but was not required.

Data extraction involved reviewer (W) identifying all empirical evidence related to psychometric and edumetric properties in line with the ACGME framework. The grading for
overall recommendation and criteria for determining the level of evidence are outlined in Table 2.

Reviewer (W) independently evaluated all included studies. Reviewer (X) blind reviewed 25% of these papers. Minor disagreements between the two reviewers were due to ambiguity in the wording of standards for validity and reliability. The descriptors for each standard were discussed by the two reviewers and consensus was reached. Once agreed, the two reviewers independently rechecked all papers and reached final consensus. After this Reviewer (W) reviewed the remaining papers holding regular meetings with Reviewer (X) to discuss any queries. Reviewer (Y) was not required during the process.

Results

Search Results and Article Overview

The initial search identified 4436 articles. PRISMA guidelines were followed in relation to protocol design and data extraction (Figure 1). Sixty-two articles met the initial study criteria after title and abstract review. Following full text review, twenty papers remained describing 14 CPATs. The characteristics of these tools are summarised in Table 3.

Fourteen of the studies were prospective cohort studies and six were retrospective in design. Seven CPATs, the Physical Therapy Clinical Performance Instrument (PT CPI 1997 and 2002 versions) (15, 16, 18, 20), Physical Therapy Manual for the Assessment of Clinical Skills (PT MACS) (11, 22), APP (14, 23, 24), Clinical Performance Assessment Form (CPAF) (25, 26), Assessment of Clinical Performance (ACP) (27) and Common Clinical Assessment Form (CAF) (28) had multiple institution involvement during testing. Four
CPATs had singular institution involvement (17, 21, 29, 30). Three CPATS, the University of Birmingham tool (UoB) (10), the Clinical Competency Evaluation Instrument (CCEVI) (12) and the Clinical Competence Evaluation Scale (CCS) (31) did not provide enough information to determine institution involvement.

All CPATs identified were similar in layout. Performance criteria (ranging from 8-53 items) were outlined for all tools. All but two used visual analogue scales, categorical rating scales or Likert scales for grading students. Two CPATs were graded by assigning a percentage for each performance criteria within the form (28) or per section of the form (29).

Validity Evidence

Content validity was described for 12 CPATs and construct validity outlined for seven (See Table 4). Criterion validity was demonstrated for one assessment tool (ECC). Two CPATs, APP and UoB, scored highest for validity evidence, both scoring Level A evidence. Both versions of the PT CPI were awarded an overall B grade for validity evidence. The remaining 10 CPATs provided insufficient information relating to validity evidence for an overall grade to be assigned.

Reliability Evidence

Evidence of Internal Consistency was available for eight tools (Table 4). One paper described test-retest reliability (13). Eight tools demonstrated evidence of inter-rater reliability (Table 4). The highest overall grading for evidence of reliability was awarded to the APP and the UoB tool (Level B) while the CAF, CPAF and CCEVI were awarded a Level C grade. There was insufficient information to grade any of the other tools.
Edumetric Evidence

All CPATs scored a Level B grade regarding their ease of use apart from one tool (ACP) which provided insufficient information to judge. The time taken to complete a student assessment using any of the reviewed CPATs was either not reported or took longer than 20 minutes; hence no CPAT was awarded an A grade for this edumetric property. Information regarding resources required to implement CPATs was inconsistent, with no tool scoring higher than a B grade (see Table 4). Nine of the 14 CPATs were awarded a C grade for evidence of ease of interpretation. No tool provided sufficient information to warrant a grade for educational impact.

Overall ACGME Grade and Summary of Evidence

No CPAT was awarded an overall Class 1 recommendation or Level A evidence. The highest grading for overall recommendation (Table 4) was Class 2 awarded to PT CPI (1997 version), APP, and UoB. No tool was awarded higher than Level C evidence. The award of a B grade or above required a CPAT to demonstrate published evidence in a minimum of two settings of all components required by the framework (validity, reliability, ease of use, resources required, ease of interpretation, educational impact).

Discussion

This systematic review identified and synthesised available evidence related to psychometric and edumetric testing of 14 CPATs used in physiotherapy practice education from 20 peer-reviewed papers. Five CPATs were developed in the USA, two in Canada, two in Ireland, and one each in United Kingdom, Australia, Japan, Malaysia and South Africa. With several hundred recognised entry level physical therapy education programmes in existence (32), this indicates a low level of publications meeting the inclusion criteria, thus highlighting the
paucity of research in this area. No Class 1 recommendation was made, and no CPAT scored higher than Level C evidence based on the framework criteria. These findings suggest limitations in the robustness of these tools. Evidence from medical education and nursing have emphasised the need for rigorous testing of CPATs highlighting that inconsistencies in these areas may affect judgments made by assessors and have implications on graduates’ readiness for practice and patient safety (3, 6).

Psychometric Evidence

Eight CPATs demonstrated internal consistency and inter-rater reliability. No reasons were provided to explain why these tests had not been carried out on the other CPATs, apart from one (16) which reported that because inter-rater reliability had been established for an earlier version of the tool, it was not necessary to repeat it although the tool had been significantly modified. Evidence of test-retest reliability was provided for one CPAT (12). Test-retest reliability is an essential property of any CPAT to ensure robust grading and ranking of students when regular observation of clinical performance is the cornerstone to deciding the final grade awarded.

Evidence of validity testing was also inconsistent; some CPATs demonstrated significant testing (APP and UoB), others provided little or no evidence (Blue MACS and RCSI tool). In several papers, information regarding the development and validity testing of the tool was only briefly described. This, together with some ambiguity in the wording of framework standards led to early discussion between the two reviewers to agree on what was acceptable.

Eleven CPATs used Likert scales, rating scales or visual analogue scales to assess student performance despite recent research identifying difficulties with these scales (33). Such
scales employ ordinal data which infers that while response categories have a ranking, the intervals between the values offered may not be assumed as equal. The problem is exacerbated when attempts are made to convert these scores to numerical grades as means and standard deviations are inappropriate for such data. Modes or medians should be employed, however this is rarely acknowledged. Experts have described this practice as providing meaningless information, encouraging competition, stress and anxiety among health professional students rather than promoting continuous progression of learning in the clinical learning environment (33, 34).

A further growing body of evidence suggests that psychometric tests alone are insufficient to assess the complexity of clinical performance, and can lead to over-emphasis on standardisation of tests, providing meaningless conversion of behaviours such as communication and team working skills to numbers (33-35). It has been suggested that a multifaceted approach to clinical performance assessment may be more robust, and may help strengthen the validity and reliability of the assessment process (5, 36). The Mini CEX and Global Rating Scales are examples of tools used in medicine and nursing which have demonstrated high levels of validity and reliability (13, 37, 38). Other clinical performance assessment adjuncts used in physiotherapy education include professional portfolios, student self-assessment and reflective essays. These have been criticised due to lack of psychometric evidence to support their sustained use in a singular context. However, medical education research has recently begun re-examining the benefits of subjective data gathered in the clinical learning environment using sensible multiple sampling. This, employed in conjunction with psychometric tests, may provide a more holistic picture of a student’s readiness to practice (9, 33, 34, 37). The physiotherapy profession could benefit from examining clinical performance assessment methods employed by other health professions.
Consideration of existing assessment methods in physiotherapy may be worth re-exploring as potential components of a clinical performance assessment toolbox rather than as sole assessment methods.

**Edumetric Evidence**

Reporting of edumetric properties was inconsistent across all CPATs. While improvements were often demonstrated in student grades from midway to the end of placement, it was impossible to attribute this to the educational impact of the CPAT used. Evidence for ease of interpretation and ease of use was variable, perhaps explained by the fact that it was not a specific aim of any of the studies. A discrepancy is apparent between the priority placed on edumetric properties by Higher Education Institutions (HEIs) during CPAT development and the importance of these properties to practice educators in the workplace. HEIs rely heavily on their clinical colleagues to facilitate and assess student learning. The practicality and feasibility of an assessment tool used in a time-constrained environment must be considered as paramount when developing an assessment tool. Findings from this review are similar to a recent review in medical education (13) which also criticised the lack of evidence for educational impact and evaluation of educational outcomes. These findings reiterate the need for greater emphasis on edumetric testing in the early stages of CPAT development.

**Collaboration**

The APP and PT CPI (1997 version) each had three associated research studies involving large numbers of participants. Recent work (38) has highlighted the significance of multiple institution involvement in psychometric testing demonstrating how validity evidence can accrue for individual assessment tools when multiple studies are involved e.g. Mini Clinical Evaluation Exercise (38), Global Rating Scale (13). Rigorous testing is necessary and
achievable in physiotherapy if greater emphasis were placed on developing and validating existing tools rather than creating new ones, and focussing efforts on larger studies. The surfeit of assessment tools that have developed across physiotherapy education has compounded this problem largely due to the development of individual tools unique to single institutions. Ongoing development and testing of existing assessment instruments has been advocated (39, 40) and should be considered by the physiotherapy profession in order to ensure more rigorous testing of assessment tools. This is especially important in the context of this study findings and the poor levels of evidence demonstrated for most CPATs.

Nationally agreed assessment tools are relatively new in physiotherapy education. Investigation of the feasibility and acceptability of the APP (originating from Australia) was explored by a physiotherapy programme in Canada (14) but not implemented there as the ACP (14, 23, 24) was developed and introduced around the same time. This suggests that assessment tools may have an inherently localised context which may make internationally acceptable and agreed assessment tools problematic. Factors complicating this may be the variation across countries in entry requirements to physiotherapy programmes, programme delivery and the final qualification obtained (32). Whether there is potential for international collaboration in this regard is unclear but WCPT may provide the impetus for initiating discussion in this area.

**Limitations of the Study**

The ACGME framework had been employed previously in medical education (13). Those authors also reported ambiguity in the wording of standards within the framework. It is possible that the current reviewers’ interpretation may have varied slightly from the original framework, but they remained consistent throughout the review process for this paper. The
reviewers were both physiotherapists who had been working in the area of physiotherapy education for over ten years; this may have facilitated consensus during the review process.

**Conclusion**

Findings from this review highlight that further discussion regarding development of CPATs within physiotherapy practice education is necessary to ensure rigour of this assessment process. A demand for objectivity, accuracy and consistency of student grading by practice educators in the midst of prioritisation of core clinical duties on a day-to-day basis is critical in order to achieve readiness for practice and patient safety assurance. Further discussion should occur in the context of evaluating other adjuncts to the clinical performance assessment process. Engagement with medicine and other allied health education providers may provide the physiotherapy profession with new perspectives on clinical performance assessment that may assist with reinforcing current processes. The WCPT may be the keystone towards enabling this discussion to occur, in particular looking towards international collaboration during CPAT development.

**Ethical Approval:** not required

**Funding:** none

**Conflict of Interest:** none
References


2. WCPT World Confederation for Physical Therapy. WCPT guideline for the clinical education component of the physical therapist professional entry-level programme. London: World Confederation of Physical Therapy, 2011.


Figure 1: PRISMA Flow Diagram of Database Search & Results

- **Identification**
  - Medline: (n=106)
  - Web of Science: (n=93)
  - Scopus: (n=756)
  - Google Scholar: (n=3479)
  - Total Citations: (n=4436)

- **Screening**
  - Duplicates removed: (n=196)
  - Total no. articles screened: (n=4330)
  - Articles Excluded: (n=4268)

- **Eligibility**
  - Full text articles assessed for eligibility: (n=62)

- **Included**
  - Articles included in review and synthesis: (n=20)
  - Reasons:
    - No psychometric data: 20
    - Not observation based: 15
    - Review articles: 4
    - Audit of tool use: 3
  - Number of clinical performance assessment tools included in papers: (n=14)
### Tables

#### Table 1: Search terms used for database search

<table>
<thead>
<tr>
<th>Competence</th>
<th>Instrument</th>
<th>Discipline</th>
<th>Learner level</th>
<th>Learning Environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical Performance*</td>
<td>Assessment</td>
<td>Physiotherap*</td>
<td>Student*</td>
<td>Practice Education</td>
</tr>
<tr>
<td>Clinical Competenc*</td>
<td>Evaluat*</td>
<td>Physical Therap*</td>
<td>Trainee*</td>
<td>Clinical Placement</td>
</tr>
<tr>
<td>Clinical Skill*</td>
<td>Apprais*</td>
<td></td>
<td>Undergraduate*</td>
<td>Practice Placement</td>
</tr>
<tr>
<td></td>
<td>Tool*</td>
<td></td>
<td></td>
<td>Clinical Education</td>
</tr>
<tr>
<td></td>
<td>Instrument*</td>
<td></td>
<td></td>
<td>Clinical Teaching</td>
</tr>
<tr>
<td></td>
<td>Measure*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Table 2: ACGME Grading for the Overall Recommendation & Criteria for Determining Level of Evidence (Swing et al., 2009)

<table>
<thead>
<tr>
<th>Grading for the Overall Recommendation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Class 1</strong></td>
<td>The assessment method is <em>recommended</em> as a core component of the program’s evaluation system.</td>
</tr>
<tr>
<td><strong>Class 2</strong></td>
<td>The assessment method <em>can be considered for use</em> as one component of the program’s evaluation system.</td>
</tr>
<tr>
<td><strong>Class 3</strong></td>
<td>The assessment method <em>can be used provisionally</em> as a component of the program’s evaluation process. Significant gaps in understanding of the assessment’s value remain, so methods in this class are best suited for investigational research.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Criteria for Determining Level of Evidence</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level A</strong></td>
<td>Published data from methodologically sound evaluation studies of the method in multiple (more than 2) settings provides strong evidence for all components of the modified utility index (reliability, validity, ease of use, resources required, ease of interpretation, and educational impact).</td>
</tr>
<tr>
<td><strong>Level B</strong></td>
<td>Published data from methodologically sound evaluation studies of the method in a minimum of two settings provides some evidence of acceptable reliability and some evidence of validity and, ease of use, and educational impact. Acceptable evidence for ease of interpretation is available for methods used to make high-stakes decisions. Available evidence for ease of use and resources required suggests that the tool is usable by many programs.</td>
</tr>
<tr>
<td><strong>Level C</strong></td>
<td>Data from methodologically sound evaluation studies of the method provide evidence of acceptable reliability, validity, or educational impact. Available evidence for ease of use and resources required suggests that the tool is usable by many programs.</td>
</tr>
<tr>
<td>Name of Tool</td>
<td>Associated Citations</td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>----------------------------------------------------------</td>
</tr>
<tr>
<td>Examination of Clinical Competence (ECC)</td>
<td>Loomis et al., 1985</td>
</tr>
<tr>
<td>Blue MACS</td>
<td>Hrachovy et al., 2004</td>
</tr>
<tr>
<td>Clinical Internship Evaluation Tool (CIET)</td>
<td>Fitzgerald et al., 2004</td>
</tr>
<tr>
<td>PT CPI (1997 version)</td>
<td>Roach et al., 2002; Adams et al., 2008; Proctor et al., 2010</td>
</tr>
<tr>
<td>Royal College of Surgeons Ireland Tool (RCSI Tool)</td>
<td>Meldrum et al., 2008</td>
</tr>
<tr>
<td>Common Clinical Assessment Form (CAF)</td>
<td>Coote et al., 2008</td>
</tr>
<tr>
<td>PT MACS</td>
<td>Stickley et al., 2004; Luedtke Hoffman et al., 2012.</td>
</tr>
<tr>
<td>Univ. Birmingham Tool (UoB)</td>
<td>Cross et al., 2010</td>
</tr>
<tr>
<td>Clinical Competence Scale (CCS)</td>
<td>Yoshino et al., 2010</td>
</tr>
<tr>
<td>Clinical Performance Assessment Form (CPAF)</td>
<td>Joseph et al., 2011; Joseph et al., 2012</td>
</tr>
<tr>
<td>PT CPI (2006 version)</td>
<td>Roach et al., 2012</td>
</tr>
<tr>
<td>Assessment of Physiotherapy Practice (APP)</td>
<td>Dalton et al., 2011; Dalton et al., 2012; Murphy et al., 2014</td>
</tr>
<tr>
<td>Canadian Physical Therapy Assessment of Clinical Performance (ACP)</td>
<td>Mori et al, 2015</td>
</tr>
<tr>
<td>Clinical Competence Evaluation Instrument (CCEVI)</td>
<td>Muhamed et al., 2015</td>
</tr>
</tbody>
</table>
Table 4: ACGME Grading for the Overall Recommendation & Level of Evidence

<table>
<thead>
<tr>
<th>Name of Tool</th>
<th>Validity Evidence</th>
<th>Overall Reliability</th>
<th>Reliability Evidence</th>
<th>Estimetric Properties</th>
<th>Overall Evidence Grade</th>
<th>Overall Recommendation Class</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Content Validity</td>
<td>Construct Validity</td>
<td>Internal Consistency</td>
<td>Inter-Rater Reliability</td>
<td>Test-Retest Reliability</td>
<td>Ease of Use</td>
</tr>
<tr>
<td>1. Evaluation of Clinical Competence (ECC)</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>N</td>
</tr>
<tr>
<td>2. Blue MAICS</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>N</td>
</tr>
<tr>
<td>3. Physical Therapist Manual for the Assessment of Clinical Skills (PT NASC)</td>
<td>YES</td>
<td>NI</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>N</td>
</tr>
<tr>
<td>5. Physical Therapist Clinical Performance Instrument (PT CPI, 2006 version)</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>N</td>
</tr>
<tr>
<td>6. Royal College Surgeons Ireland (RCSI) Tool</td>
<td>NI</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>N</td>
</tr>
<tr>
<td>7. Common Clinical Assessment Form (CCAF)</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>N</td>
</tr>
<tr>
<td>8. University of Birmingham Tool (ubPT)</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>B</td>
</tr>
<tr>
<td>9. Assessment of Physiotherapy Practice (APP)</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
<td>B</td>
</tr>
<tr>
<td>10. Clinical Performance Assessment Form (CPAF)</td>
<td>YES</td>
<td>NO</td>
<td>NI</td>
<td>YES</td>
<td>YES</td>
<td>NO</td>
</tr>
<tr>
<td>11. Clinical Competence Scale (CCS)</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>YES</td>
<td>NI</td>
<td>NO</td>
</tr>
<tr>
<td>12. Clinical Internship Evaluation Tool (CET)</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>N</td>
</tr>
<tr>
<td>13. Canadian PT Assessment of Clinical Performance (CPAP)</td>
<td>YES</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>NO</td>
<td>N</td>
</tr>
</tbody>
</table>

NI: Not enough information is given to judge.