Journal of Research and Curriculum Development, Vol.14, Issue.2, January – June, 2024
ISSN: 2730-387X (Online) https://so03.tci-thaijo.org/index.php/jrcd/index

Inter-rater Reliability of a Clinical Practice Assessment Rubric for Physical Therapy Students: Mahidol University

Rujiret Upiriyasakul[1], Sivaporn Limpaninlachat[2]*, Pavinee Harutaichun[3], Jenjira Thanakamchokchai[4], Kotchapatch Keawmen[5], Sirikarn Somprasong[6]
[1-6]Faculty of Physical Therapy, Mahidol University, Nakhon Pathom, Thailand 73170
*Corresponding author e-mail: sivaporn.lim@mahidol.ac.th

Received: January 4, 2024
Revised: June 15, 2024
Accepted: September 10, 2024

Abstract: The rubric for clinical practice training is content-accurate, however a student can be assessed more than one clinical instructor by occasion. This may be affected to achieve the learning objectives and outcomes. It is thus this study aim to investigate inter-rater reliability of a clinical practice assessment rubric. The researcher compiled data of learners who were evaluated in competency by more than one clinical instructor during the same period, in the academic years 2020- 2022. The assessment rubric includes three aspects of competency: 1 item in the knowledge aspect, 10 items in the skill aspect, and 9 items in the professional behavior aspect, for statistical analysis and to consider the Intraclass Correlation Coefficient (ICC 2,2). From a dataset of 158 scores, good inter-rater reliability was observed in questions related to patient care standards (ICC = 0.91), self-improvement in knowledge (ICC = 0.83), and effective communication (ICC = 0.79). Moderate reliability was found in most skill-related items, such as hypothesis formulation (ICC = 0.73) and physical examination skills (ICC = 0.56). However, four professional behavior items and three skill-related items exhibited low reliability, indicating a need for further refinement. The agreement on conditions of assessment and the process of training evaluators needs to be reviewed to guarantee the rubric's reliability among clinical instructors. Additionally, the operational definition should be refined to provide reliable and objective competency assessments of physical therapy students, leading to more effective and consistent evaluations.

Keywords: Clinical practice training, Rubric assessment, Inter-rater reliability, Physical therapy

Introduction

Physical therapy education institutions incorporate clinical practice into their curriculum to build competencies according to the profession. As defined in The Physical Therapy Profession Act, B.E. 2547 (The Physical Therapy Profession Act, B.E. 2547, 2004), physical therapy involves examining, diagnosing, and treating physical impairments caused by disease or abnormal movement, as well as preventing, correcting, and rehabilitating physical deterioration and disabilities, including promoting physical and mental health through physical therapy methods or designated instruments and equipment.

Physical therapy students must develop methods for assessment, diagnosis, treatment planning, and treatment that reference foundational and applied knowledge with precision and comprehensiveness. They must also be able to record diagnostic results and treatment plans to aid in patient care. These competencies are assessed in the third and fourth years of study, supplementing lectures, physical therapy practice, and case study learning.

Measurement and assessment are crucial for reflecting whether students achieve the learning objectives and outcomes. Rubric assessment, a competency-based approach in medical education, focuses on outcomes. Rubrics assess real-world conditions, allowing for both quantitative and qualitative interpretations of performance against set criteria. This assessment provides feedback to both students and educators for development (Y. Malini & Heidi, 2010). Rubrics are reliable and widely used tools, clearly defining each performance level (Andrade, 2005; Dawson, 2017; Jonson, 2007; Kohn, 1995; Malini & Heidi, 2010). They range from highest competency, descending through a 5-level scale, covering two levels above and two below the standard (Li & Qian, 2021). Rubrics have been successfully

applied in medical (Smith et al., 2016), dental (O'Donnell et al., 2011), and nursing education (Julie J & Annette S., 2009), proving accurate and precise in student assessment.

Mahidol University's Physical Therapy Faculty has been developing and implementing rubric assessments since 2019 to evaluate clinical teaching, reflect on processes, and enhance the effectiveness of educational management in health-related disciplines, as suggested by studies in medicine (Smith et al., 2016), dentistry (O'Donnell et al., 2011), and others. In use of clinical practice one student had occasion to be evaluated by at least one clinical instructor. Thus, the rubric assessment for physical therapy students was studied only content validity in knowledge, skills, and professional behavior (Kongoun et al., 2019). It lacked reliability data, which is critical as students are evaluated by more than one assessor during the same training period. This may be affected to achieve the learning objectives and outcomes. It is aimed to investigate inter-rater reliability of a clinical practice assessment rubric.

This prompted the study to test the rubric's efficacy for physical therapy student clinical practice assessments.

Research methodology

This retrospective study involved 158 physical therapy students from Mahidol University, who were assessed using a rubric for clinical practice in Clinical Practice Courses 2, 3, and 4 during the academic years 2020, 2021, and 2022. These students had received evaluations from clinical instructors in more than one of these courses. The clinical instructors are certified physical therapist with more than two years of experienced work in physical therapy.

In this retrospective study design, the researcher obtained ethical approval from the Mahidol University Human Research Ethics Committee, with the project code MU-CIRB 2023/265.2408. This approval allowed for the collection of historical data regarding the students' scores in the aforementioned clinical practice courses. The researcher sought data for students registered in these courses who met specific criteria: those evaluated by clinical instructors more than once during the same period and who received their grades within the normal timeframe without course withdrawal or interruption during the semester. The researcher organized the student codes and performance scores, encompassing competencies in knowledge, skills, and professional behavior. The assessment contained 20 questions: 1 on knowledge, 10 on skills, and 9 on professional behavior, scored on a 5-level scale from significantly above the standard (5 points) to significantly below the standard (1 point). The scores were adjusted based on the weight of percentages in each question, as detailed in Table 1, for subsequent statistical analysis.

Table 1 Competencies in knowledge, skills, and professional behavior (Question, Q1-Q20).

Competencies	Questions
Knowledge	
Question 1	Ability to apply basic knowledge and physical therapy principles (30%).
Skills	
Question 2	Capability to conduct patient history interviews and gather patient data from medical records (4%).
Question 3	Skill in formulating reasonable hypotheses (4%).

Table 1 Continue

Question 4	Ability to plan physical examinations (4%).
Question 5	Proficiency in physical examination skills (4%).
Question 6	Capacity to interpret physical examination results, diagnose physical therapy conditions, and prognosticate diseases (4%).
Question 7	Competence in summarizing physical therapy problems based on the ICF model (4%).
Question 8	Ability to set goals and plan treatments (4%).
Question 9	Skills in managing physical therapy (4%).
Question 10	Aptitude for teaching and providing guidance to patients and/or caregivers (4%).
Question 11	Ability to write reports or record patient data in medical files (4%).
Professional behavior	
Question 12	Behavior that demonstrates altruism and kindness towards others (3%).
Question 13	Behavior that shows caution for safety and reduces risks for patients, oneself, and others (4%).
Question 14	Professional conduct towards patients, colleagues, and other staff, including appropriate dressing and behavior (3%).
Question 15	Skills in communicating with patients and various personnel using appropriate verbal and non-verbal language (3%).
Question 16	Responsibility in punctuality and timely task completion during internships (3%).
Question 17	Providing patient care according to professional standards and respecting patient rights (3%).
Question 18	Adherence to the rules and regulations of the healthcare facility (3%).
Question 19	Continuous development of personal knowledge and treatment skills (4%).
Question 20	Demonstration of self-improvement in behavior during internships (4%).

Statistical Analysis

The sample size calculation will use Bonett's 2023 formula (9) as shown in the following equation, where $\alpha = 0.05$, $\beta = 0.1$, $\rho a = 0.3$, $\rho b = 0.7$, $\delta = 2.3$, with ρa and ρb ranging from -1 to 1. Therefore, at least 68 participants will be included in this study. However, the sample size will be determined based on the number of registrations in the Clinical Practice Course, Bachelor of Science in Physical Therapy program.

The researcher used the Intraclass Correlation Coefficient (ICC 2,2) to measure the consistency of scores between evaluators. An ICC of less than 0.5 indicates low reliability, between 0.5 and 0.75 indicates moderate reliability, between 0.75 and 0.9 indicates good reliability, and above 0.9 indicates excellent reliability (Koo & Li, 2016).

Research results

The study found that of the 158 student codes, several assessment items had low reliability, including 3 skill-related items and 4 professional behavior items. Additionally, 10 items had moderate reliability, including 1 knowledge-related item, 7 skill-related items, and 2 professional behavior items. Finally, 3 items related to professional behavior had good reliability. The complete study results are presented in Table 2.

Table 2 shows the level of agreement between different assessors for each assessment question based on competencies.

Table 2 Shows the level of agreement between different assessors for each assessment question based on competencies.

Competencies	Questions	Assesssor 1	Assesssor 2	ICC (95%CI)
		Mean (SD)	Mean (SD)	
Professional behavior	Q17 Providing patient care according to professional standards and respecting patient rights	3.28 (0.50)	3.25 (0.50)	0.91 (0.870-0.930)
Professional behavior	Q19 Demonstration of self-improvement in knowledge and skill during internships	3.34 (0.60)	3.31 (0.68)	0.83 (0.76-0.87)
Professional behavior	Q15 Skills in communicating with patients and various personnel using appropriate verbal and non-verbal language	3.16 (0.54)	3.12 (0.63)	0.79 (0.72-0.85)
Skills	Q3 Skill in formulating reasonable hypotheses	3.17 (0.59)	3.18 (0.68)	0.73 (0.63-0.80)
Professional behavior	Q20 Demonstration of self-improvement in behavior during internships	3.37 (0.62)	3.39 (0.62)	0.73 (0.63-0.80)
Knowledge	Q1 Ability to apply basic knowledge and physical therapy principles	23.39 (3.84)	23.32 (5.07)	0.73 (0.63-0.80)
Professional behavior	Q16 Responsibility in punctuality and timely task completion during internships	2.89 (0.33)	2.81 (0.40)	0.68 (0.55-0.76)
Skills	Q6 Capacity to interpret physical examination results, diagnose physical therapy conditions, and prognosticate diseases	3.17 (0.54)	3.21 (0.65)	0.63 (0.49-0.73)
Skills	Q5 Proficiency in physical examination skills	3.21 (0.53)	3.23 (0.61)	0.56 (0.40-0.68)
Skills	Q8 Ability to set goals and plan treatments	3.33 (0.57)	3.29 (0.58)	0.54 (0.37-0.66)
Skills	Q7 Competence in summarizing physical therapy problems based on the ICF model	3.34 (0.56)	3.37 (0.58)	0.54 (0.37-0.66)
Skills	Q9 Skills in managing physical therapy	3.35 (0.51)	3.23 (0.61)	0.52 (0.34-0.65)
Skills	Q4 Ability to plan physical examinations	3.19 (0.56)	3.24 (0.57)	0.52 (0.35-0.65)

Table 2 Continue

Competencies	Questions	Assesssor 1	Assesssor 2	ICC (95%CI)
		Mean (SD)	Mean (SD)	
Skills	Q10 Aptitude for teaching and providing guidance to patients and/or caregivers	3.46 (0.57)	3.39 (0.63)	0.44 (0.24-0.60)
Skills	Q11 Ability to write reports or record patient data in medical files	3.33 (0.55)	3.31 (0.61)	0.45 (0.24-0.59)
Skills	Q2 Capability to conduct patient history interviews and gather patient data from medical records	3.31 (0.55)	3.30 (5.29)	0.43 (0.22-0.59)
Professional behavior	Q12 Behavior that demonstrates altruism and kindness towards others	2.84 (0.30)	2.85 (0.32)	0.32 (0.06-0.50)
Professional behavior	Q14 Professional conduct towards patients, colleagues, and other staff, including appropriate dressing and behavior	2.90 (0.27)	2.84 (0.38)	0.21 (-0.08-0.42)
Professional behavior	Q18 Adherence to the rules and regulations of the healthcare facility	2.98 (0.11)	2.94 (0.19)	0.07 (-0.27-0.32)
Professional behavior	Q13 Behavior that shows caution for safety and reduces risks for patients, oneself, and others	3.79 (0.38)	3.82 (0.39)	0.03 (-0.33-0.29)

Abbreviations: ICC = intra-class correlation coefficients, CI = confidence interval
Color: white bars indicate good reliability, light bars indicate moderate reliability, and dark bars indicate low reliability

Discussion

The results of this study demonstrate several strengths of the rubrics created for clinical practice assessments in physical therapy students at Mahidol University. One key strength is the good reliability observed in certain competencies, particularly in the areas of professional behavior. Specifically, the questions related to providing patient care according to professional standards and respecting patient rights (ICC = 0.91), demonstrating self-improvement in knowledge and skill during internships (ICC = 0.83), and effective communication with patients and various personnel using appropriate verbal and non-verbal language (ICC = 0.79). Additionally, the moderate reliability observed in most skill-related items, such as formulating reasonable hypotheses (ICC = 0.73), planning examinations (ICC = 0.52), proficiency in physical examination skills (ICC = 0.56), and interpreting

examination results (ICC = 0.63), highlights the rubric's utility in evaluating technical skills essential for clinical practice.

However, four items of professional behavior and three related skill items were low reliability. These illustrated either skill- related or professional behavior- related have no consistency value of reliability. The results may occur from deficit operational definition. In addition, ranking score is rated by proportional decision such as five-performance; good performance completely, good permeances, performance adjustably, fair performance, poor performance, five occasion; 80-100%, 61-80%, 51-60%, 21-50%, less than 20%. The previous studies showed the operational definition is clarified clearly affected to prevent confusion and ambiguity among evaluators (Malini & Heidi, 2010). Similarly, each ranking score should include defined sub-items or checklists to improve inter-rater reliability. For instance, "Question 2: Capability to conduct patient history interviews and gather patient data from medical records" might be refined to include a required performance checklist. Because physical therapy skills can be applied to patients with various conditions, including orthopedics, pediatrics, neurology, cardiovascular, and pulmonary, the performance of students may be assessed differently by clinical instructors.

Consistent understanding and application of evaluation criteria are also key to enhancing the reliability of the assessment system. The process of training and enhancing evaluator performance needs to be revisited to develop more reliable assessment criteria (Jonson, 2007).

Conclusion

Therefore, the application of the Clinical Practice Assessment Rubric for physical therapy students should include crucial training for all clinical instructors to ensure that the ranking scores align with the required performance of the course. For further study, the operational definition should be listed and aligned with the goal standard assessment tool. This enhancement will ensure that the assessment of physical therapy students' competencies is more reliable and closely aligned with the desired objective outcomes, ultimately leading to more effective and consistent evaluations.

References

Andrade, H. G. (2005). Teaching with rubrics: The good, the bad, and the ugly. *College Teaching, 53*(1), 27–31.

Dawson, P. (2017). Assessment rubrics: towards clearer and more replicable design, research and practice. *Assessment & Evaluation in Higher Education, 42*(3), 347–360.

Jonson, A. S. G. (2007). The use of scoring rubrics: reliability, validity and educational consequences. *Educational Research Review*, 2, 130–144.

Julie, J. I., & Annette, S. S. (2009). Rubrics for clinical evaluation: Objectifying the subjective experience. *Nurse Education in Practice, 9*(2), 134–140.

Kohn, K. A. M. (1995). Grading practices used by clinical instructors in physical therapy in the United States. *Physiotherapy Theory and Practice, 11*(3), 175–181.

Kongoun, S., Sermpon, N., Suraprapapich, P., Nuntapornsak, A., Klomjai, W., & Vongsirinavarat, M. (2019). The Development of the Clinical Practice Rubric Assessment Form for Undergraduate Physical Therapy Students at Mahidol University. *Journal of Research Methodology, 20*(1).

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine, 15*(2), 155–163.

Li, J., & Qian, W. (2021). Development and validation of a rating scale for summarization as an integrated task. *Asian-Pacific Journal of Second and Foreign Language Education.*

Malini, R. Y., & Heidi, A. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education, 35*(4), 435–448.

O'Donnell, J. A., Oakley, M., Haney, S., O'Neill, P., & Taylor, D. (2011). Rubrics 101: A Primer for Rubric Development in Dental Education. *Journal of Dental Education, 75*(9), 1163–1175.

Smith, K., Simon, G., & Austin, M. (2016). The Use of Rubrics in the Clinical Evaluation of Podiatric Medical Students. *Journal of the American Podiatric Medical Association, 106*(1).

The Physical Therapy Profession Act, B.E. 2547. (2004).