AJSTME, Volume. 12, Issue 1. 126-132, January, 2026. https://www.ajstme.com.ng
ISSN: 2251-0141 (Print), 2971-6233 (Online)

Reliability Evidence for the Psychomotor Skills Assessment in Physics (PSAP): Inter-Rater Agreement, Stability, and Inter-Correlation Indices

[1]Nwankwo, Ugonna MaryRose, [2]U. N. Akanwa & [2]I. O. Eluwa
[1]Demonstration Secondary school Umudike, Abia State
[2]Science Education, Michael Okpara University of Agriculture Umudike
Corresponding Author: ugoonwankwo@yahoo.com

Abstract

This study established reliability evidence for the Psychomotor Skills Assessment in Physics (PSAP), a performance-based observational rating scale designed to assess senior secondary school students' psychomotor skills in Physics practical work. The study was part of a larger instrumentation study conducted in South-East Nigeria. Data were generated from 384 Senior Secondary School III Physics students using PSAP, which consists of 50 items distributed across five subscales: Basic Fundamental Movement, Perceptual Ability, Physical Ability, Skilled Movement, and Non-Discursive Communication. Inter-rater reliability was estimated using the Intraclass Correlation Coefficient (ICC), stability was determined using Pearson product-moment correlation, and inter-correlation evidence was examined through Pearson inter-correlation ranges. The findings showed that the overall instrument had a high inter-rater reliability coefficient of .88, with subscale ICC values ranging from .84 to .92. The coefficient of stability for the overall instrument was .86, while subscale stability coefficients ranged from .74 to .96. In addition, the inter-correlation matrix showed moderate to strong positive relationships across PSAP items and subscales, with coefficients ranging from .52 to .89. These findings indicate that PSAP provides dependable scores across raters, across time, and across its internal dimensions. The study concludes that PSAP is a reliable instrument for assessing students' psychomotor skills in Physics practical activities and recommends its use in school-based practical assessment.

Keywords: Inter-rater reliability, Psychomotor assessment, Physics practical, reliability evidence, stability coefficient

Introduction

Practical work is an essential aspect of Physics education, as it enables students to convert theoretical knowledge into observable actions through the manipulation of equipment, measurement, observation, and experimental processes. Engaging in physics practical activities requires learners not only to understand scientific principles but also to exhibit proficiency in carrying out tasks with accuracy, safety, and a systematic approach. This is consistent with the long-standing view that effective science learning is not complete until learners can connect conceptual understanding with direct experimentation and manipulation of materials. Deacon and Hajek (2019) and Prades and Espinar (2020) emphasized that practical work helps students verify principles, correct misconceptions, and link theory with real-life scientific activity. For this reason, assessment in Physics should extend beyond the cognitive domain to include the psychomotor domain, which captures learners' practical and manipulative abilities.

The psychomotor domain occupies a central place in science learning because it concerns the development of coordinated movement, manual skill, control, accuracy, and technical performance. In the context of education, psychomotor learning involves the physical execution of tasks that reflect the learner's ability to translate knowledge into purposeful action. Hamid and Baharom (2020) and Baharom et al. (2021) noted that psychomotor skills are usually developed in laboratory and workshop settings where students interact directly with tools, materials, and equipment. In Physics, such interaction is especially important because many school-based practical activities require learners to set up apparatus, align instruments, measure physical quantities, record data, and adjust procedures with precision. Thus, psychomotor competence is not peripheral to Physics education; it is one of the conditions for meaningful engagement with the subject.

The importance of psychomotor assessment is also supported by the goals of science education in Nigeria. The National Policy on Education emphasizes the acquisition of appropriate skills, scientific attitudes, and competencies necessary for technological development and productive living (Federal Republic of Nigeria, 2013, 2014). Similarly, the Nigerian Educational Research and Development Council physics curriculum stresses the acquisition of relevant scientific skills and attitudes as preparation for technological application and further study (NERDC, 2012). These policy and curriculum provisions imply that learners should not only know Physics concepts but should also be able to perform practical tasks that demonstrate procedural competence. Yet, as your main work rightly shows, school assessment practices often focus more on theoretical knowledge than on the direct evaluation of psychomotor skills.

Psychomotor assessment involves the evaluation of students' ability to apply knowledge through coordinated physical action and task performance. In educational practice, this includes such behaviours as handling equipment correctly, setting up apparatus, taking measurements, observing carefully, and carrying out procedures in the proper sequence. Hidayah (2017) explained that psychomotor assessment requires direct observation of learners as they prepare materials, arrange equipment, carry out procedures, and interpret outcomes. Similarly, Begam and Tholappan (2018) noted that psychomotor assessment becomes more meaningful when students are observed while performing actual practical tasks rather than when they are assessed only through written responses. In the same vein, Salim et al. (2022) argued that practical skill assessment in laboratory settings should be based on observable performance behaviours because such behaviours reveal whether students have genuinely acquired manipulative competence.

In Physics education, the need for psychomotor assessment is especially important because many practical competencies cannot be validly inferred from written reports alone. Students may provide correct theoretical explanations or well-written laboratory records without demonstrating actual skill in apparatus manipulation or experimental execution. Chabalengula et al. (2009) emphasized that direct observation during laboratory work provides a more authentic basis for assessing students' science process skills than written products alone. Abraham, Reiss, and Sharpe (2023) also argued that the assessment of practical work in science should reflect what students do in the laboratory, not merely what they later write about it. This makes performance-based instruments particularly relevant for evaluating students' practical competence in Physics.

A useful theoretical basis for understanding psychomotor performance comes from psychomotor taxonomies such as those of Simpson (1972), Dave (1975), and Harrow (1972). Among these, Harrow's taxonomy is especially relevant because it describes psychomotor learning as progressing from basic movement to more refined and complex forms of skilled performance. The taxonomy includes such levels as basic fundamental movement, perceptual abilities, physical abilities, skilled movements, and non-discursive communication. These categories are closely related to the demands of Physics practical work, where students must coordinate hand and eye movements, manipulate apparatus, exercise control and balance, and demonstrate increasing fluency in experimental procedures. By organizing assessment around these dimensions, an instrument such as PSAP can provide a more comprehensive evaluation of students' practical capabilities.

However, the usefulness of any assessment instrument depends not only on its content and structure but also on the strength of its psychometric properties. Among these properties, reliability is central. Reliability refers to the degree to which an instrument produces consistent and dependable results under appropriate conditions (Nworgu, 2015; Heale & Twycross, 2015). A reliable instrument reduces measurement error and increases confidence that observed scores reflect actual differences in students' performance rather than inconsistencies in the instrument or scoring process. In the context of psychomotor assessment, reliability is particularly critical because scores are often derived from observation and professional judgment, which may introduce rater variability if the instrument is not carefully developed and standardized.

Reliability evidence can be examined from different but complementary perspectives. One important form is inter-rater reliability, which refers to the degree of agreement among different raters who assess the same performance. This is especially important for performance-based and observational instruments where scores may be influenced by the subjective judgment of assessors. When raters interpret criteria differently, the usefulness of the instrument is weakened. Onunkwo (2015) and Nworgu (2015) both emphasized that scorer reliability is necessary for instruments involving observational ratings, checklists, and essay-type judgments. In the present context, a psychomotor assessment instrument in Physics must show that different raters can apply it consistently to the same student performance.

A second important form of reliability is stability, often estimated through a test–retest approach. Stability refers to the extent to which an instrument yields consistent scores across time when administered under comparable conditions. Vilagut (2024) defined test–retest reliability as the degree to which test scores remain unchanged when measuring a stable individual characteristic on different occasions. De Souza et al. (2017) also explained that an instrument is dependable when repeated measurement produces sufficiently similar results. In psychomotor assessment, evidence of stability is important because it shows that the instrument is not producing erratic scores from one administration to another. This is particularly relevant where the assessment is intended for classroom use, research, or monitoring of student progress

A third source of reliability evidence is the inter-correlation pattern among items and subscales. Inter-correlation analysis provides information about the internal relatedness of the instrument's dimensions and the extent to which items function together in a coherent way. When items correlate too weakly, the instrument may lack unity; when they correlate too highly, item redundancy may be indicated. Therefore, examining the inter-correlation matrix of PSAP and its subscales helps determine whether the instrument reflects a coherent but non-duplicative structure. This type of evidence is important in multidimensional observational instruments, especially those designed around theoretically distinct but related aspects of psychomotor performance.

Empirical studies reviewed in the main work also support the need for reliability-focused investigation of psychomotor instruments. Yusuf and Afolabi (2018), for example, developed an instrument for assessing students' laboratory skills in Physics and reported a high reliability coefficient, indicating that structured observation can yield dependable data in school science settings. Eze and Agwagah (2017) developed and validated a test for measuring Physics laboratory manipulative skills and demonstrated that psychomotor performance in Physics can be assessed through carefully designed practical instruments. Nahadi, Firman, and Yulina (2016) developed a performance assessment instrument for assessing senior secondary students' psychomotor competence and reported acceptable evidence of validity. Although their study was situated in chemistry, it reinforces the broader point that performance-based psychomotor assessment can be systematically developed and tested.

Further empirical support comes from technical and vocational education studies reviewed in your thesis. Asukwo et al. (2019) developed and validated a psychomotor skills assessment guide in electronics work and reported acceptable internal consistency, inter-rater reliability, and correlation over time. Hassan, Musa, and Usman (2020) similarly developed an instrument for assessing students' practical skills in electronics works and obtained reliability coefficients that supported classroom use. Hassan, Deba, and Saulawa (2022) also reported that a standardized instrument for assessing practical skills in carpentry and joinery showed strong reliability. Mbaga, Yusuf, and Asukwo (2024) extended this pattern by developing a psychomotor skill assessment tool for electrical installation and maintenance works and reporting high reliability and acceptable inter-rater agreement. Although these studies were conducted outside Physics, they provide important empirical support for the broader argument that psychomotor assessment instruments can and should be supported by multiple forms of reliability evidence.

Despite these contributions, there remains a gap in the area of senior secondary school Physics. Much of the existing research in Physics education has focused on cognitive achievement, interest, attitude, conceptual understanding, or the effects of instructional strategies, while relatively less attention has been devoted to the systematic assessment of students' psychomotor performance using standardized instruments. Where practical work is assessed, teachers often rely on informal observation or written reports rather than structured performance rating. This gap is especially important in a subject like Physics, where the ability to carry out practical procedures is central to meaningful learning and to preparation for higher study and technological application.

It was in response to this need that the Psychomotor Skills Assessment in Physics (PSAP) was developed. The PSAP is a structured observational rating scale designed to assess students' psychomotor skills in Physics practical work across five subscales: Basic Fundamental Movement, Perceptual Ability, Physical Ability, Skilled Movement, and Non-Discursive Communication. Since the instrument is intended for direct classroom and laboratory use, it is necessary to establish whether it produces dependable results across raters, across time, and across its internal dimensions.

This study therefore investigated the reliability evidence for PSAP by examining its inter-rater reliability, stability indices, and inter-correlation matrix. Establishing these reliability properties is essential for supporting the use of PSAP as a dependable instrument for assessing students' psychomotor performance in Physics practical activities.

Research Questions

The following research questions guided the study:

1. What are the inter-rater reliability coefficients of PSAP and its subscales?
2. What are the stability indices of PSAP and its subscales?
3. What is the inter-correlation matrix of PSAP and its subscales?

Methods

The study adopted an instrumentation research design. The article was extracted from a broader study that developed and validated the Psychomotor Skills Assessment in Physics (PSAP) for senior secondary school students. Instrumentation research design according to Cronbach and Meehl (1955) and Kalkbrenner (2021) is a systematic process aimed at creating or refining research tools and procedures for accurate valid and reliable data collection in a given research context. This research design is considered appropriate for this study since it aims at developing a psychomotor skill assessment tool in physics for senor secondary school students. The study was carried out in South-East Nigeria. Three states - Abia, Enugu, and Imo - were selected from the zone for the study. The choice of the three states was guided by considerations of manageability, accessibility, and resource constraints, without compromising the generalizability of the findings. The population consisted of Senior Secondary School III Physics students in the selected states. A sample of 384 students was selected through multistage sampling from 12 secondary schools. The students participated in Physics practical activities and were rated with the PSAP. Data were generated using the Psychomotor Skills Assessment in Physics (PSAP), a 50-item observational rating scale distributed across five subscales: Basic Fundamental Movement, Perceptual Ability, Physical Ability, Skilled Movement, and Non-Discursive Communication. Each subscale contained 10 items scored on a five-point rating scale during students' practical performance.

Inter-rater reliability was estimated using the Intraclass Correlation Coefficient (ICC) based on a two-way random-effects model with absolute agreement. Stability was estimated using Pearson product-moment correlation coefficients. Inter-correlation evidence was examined using Pearson inter-correlation ranges among PSAP items and subscales.

Results

Table 1. Inter-rater reliability of PSAP and its subscales

PSAP and Subscale	ICC	95% CI Lower	95% CI Upper	Interpretation
Basic Fundamental Movement	0.92	0.864	0.984	Very high
Perceptual Ability	0.85	0.843	0.876	Very high
Physical Ability	0.87	0.823	0.934	Very high
Skilled Movement	0.84	0.832	0.862	Very high
Non-Discursive Communication	0.91	0.842	0.986	Very high
Overall Instrument	0.88	0.840	0.928	Very high

Table 1 shows that the PSAP demonstrated high inter-rater reliability across all five subscales. The subscale ICC values ranged from 0.84 to 0.92, while the overall ICC for the instrument was 0.88. These coefficients indicate a high level of agreement among raters in scoring students' psychomotor performance in Physics practical activities.

Table 2. Stability indices of PSAP and its subscales

PSAP and Subscale	Number of Items	Stability (rxy)	Coefficient	Interpretation
Basic Fundamental Movement	10	0.96		Very high
Perceptual Ability	10	0.90		Very high
Physical Ability	10	0.85		High
Skilled Movement	10	0.89		Very high
Non-Discursive Communication	10	0.74		Acceptable
Overall Instrument	50	0.86		High

Table 2 indicates that PSAP was stable over time. The subscale stability coefficients ranged from 0.74 to 0.96, while the overall stability coefficient for the instrument was 0.86. These values show that the PSAP produced dependable scores across administrations.

Table 3. Summary of inter-correlation evidence for PSAP and its subscales

PSAP and Subscale	Correlation Range (r)	Interpretation
Basic Fundamental Movement	0.57-0.87	Moderate to strong
Perceptual Ability	0.62-0.88	Moderate to strong
Physical Ability	0.58-0.89	Moderate to strong
Skilled Movement	0.59-0.86	Moderate to strong
Non-Discursive Communication	0.52-0.87	Moderate to strong
Overall PSAP	0.52-0.89	Moderate to strong

Table 3 summarizes the inter-correlation evidence for PSAP. The coefficients ranged from 0.52 to 0.89 across items and subscales. No coefficient fell below 0.30 or exceeded 0.90, indicating that the items and subscales were sufficiently related without suggesting redundancy.

Discussion

The first finding of the study showed that PSAP demonstrated high inter-rater reliability across all subscales and for the overall instrument. The overall ICC of 0.88 indicates that the instrument can be scored consistently by different raters. This is important because psychomotor assessment in Physics depends heavily on observation and professional judgment during practical activities. When an instrument yields similar ratings across assessors, confidence in its scoring process is strengthened. This finding agrees with Asukwo et al. (2019), who reported acceptable inter-rater reliability in a psychomotor skills assessment guide, and with Umar and Hassan (2023), who reported strong inter-rater agreement in a hands-on task assessment for Physics practical performance. The result suggests that PSAP has sufficient scorer consistency for school-based practical assessment.

The second finding showed that PSAP possessed high stability over time. The overall stability coefficient of 0.86 indicates that the instrument produced dependable scores across repeated administrations, while the subscale coefficients also remained within acceptable to very high ranges. This means that the instrument was not overly sensitive to temporary fluctuations and could provide stable estimates of students' psychomotor performance. This finding is consistent with De Souza et al. (2017), who emphasized that repeated measurement should yield sufficiently similar results if an instrument is dependable. It also agrees with Yusuf and Afolabi (2018), who reported strong reliability evidence in an instrument designed to assess laboratory skills in Physics.

A third source of reliability evidence is the inter-correlation pattern among items and subscales. Inter-correlation analysis provides information about the internal relatedness of the instrument's dimensions and the extent to which items function together in a coherent way. Madadizadeh and Bahariniya (2025) explained that internal consistency is typically assessed from the correlations among different items on the same test or subscale and shows whether items intended to measure the same construct produce similar scores. When items correlate too weakly, the instrument may lack unity; when they correlate too highly, item redundancy may be indicated. Therefore, examining the inter-correlation matrix of PSAP and its subscales helps determine whether the instrument reflects a coherent but non-duplicative structure. This type of evidence is important in multidimensional observational instruments, especially those designed around theoretically distinct but related aspects of psychomotor performance.

Conclusion

The study established three complementary forms of reliability evidence for the Psychomotor Skills Assessment in Physics (PSAP). First, the instrument showed high inter-rater agreement across all subscales and in the overall score. Second, it demonstrated strong stability over time. Third, it showed moderate to strong positive inter-correlations across items and subscales, indicating internal coherence without problematic redundancy. Taken together, these findings support the use of PSAP as a dependable instrument for assessing senior secondary school students' psychomotor skills in Physics practical activities.

Recommendations

1. Physics teachers and school-based assessors should use the PSAP with confidence in practical Physics evaluation because the instrument demonstrated high inter-rater reliability across raters.
2. The PSAP should be adopted by physics teachers for repeated classroom and school-based practical assessment because the instrument showed stable performance across measurement occasions.
3. The five PSAP subscales should be retained together in practical assessment because the inter-correlation evidence showed that they function as related but non-redundant dimensions of students' psychomotor skills in Physics.

References

Abraham, I., Reiss, M. J., & Sharpe, R. M. (2023). The assessment of practical work in school science. *Studies in Science Education*, *49*(2), 209–251.

Asukwo, A. E., Chatramada, V., Patience, M. D., & Mohammed, U. M. (2019). Development and validation of psychomotor skills assessment guide for repairs of television receivers in electronics work for technical college in Akwa Ibom State, Nigeria. *International Journal of Educational Benchmark*, *13*(2), 1–18.

Baharom, S., Khoiry, M. A., Hamid, R., Mutalib, A. A., & Hamzah, N. (2021). Assessment of psychomotor domain in a problem-based concrete laboratory. *Journal of Engineering Science and Technology*, 1–10.

Begam, A. A. A., & Tholappan, A. (2018). Psychomotor domain of Bloom's taxonomy in teacher education. *Shanlax International Journal of Education*, *6*(3), 11–14. https://doi.org/10.17499/sijed.2018.6.3.300823

Chabalengula, V. M., Mumba, F., Hunter, W. F. J., & Wilson, E. (2009). A model for assessing students' science process skills during science lab work. *Problems of Education in the 21st Century*, *11*, 28–36.

Dave, R. H. (1975). *Developing and writing behavioural objectives*. Educational Innovation Press.

De Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços de Saúde*, *26*(3), 649–659. https://doi.org/10.5123/S1679-49742017000300022

Deacon, C., & Hajek, A. (2019). Student perceptions of the value of physics laboratories. *International Journal of Science Education*, *33*(7), 943–977. https://doi.org/10.1080/09500693.2010.481682

Eze, J. U., & Agwagah, U. N. V. (2017). Developing and validating a test for measuring physics laboratory manipulative skills. *Journal of the Science Teachers Association of Nigeria*, *52*(1), 115–124.

Federal Republic of Nigeria. (2013). *National policy on education*. NERDC Press.

Federal Republic of Nigeria. (2014). *National policy on education*. NERDC Press.

Hamid, R., & Baharom, S. (2020). Monitoring of laboratory work problem-based project implementation. In *2013 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)* (pp. 585–588). IEEE.

Harrow, A. J. (1972). *A taxonomy of the psychomotor domain: A guide for developing behavioral objectives*. David McKay.

Hidayah, M. (2017). Pengembangan instrumen penilaian psikomotor pada outdoor practicum biologi SMA. *Didaktika Biologi: Jurnal Penelitian Pendidikan Biologi*, *1*(2), 143–148.

Hassan, A. A., Musa, H. Y., & Usman, M. Y. (2020). Development and validation of an instrument for assessing students' practical skills in electronics works trade in technical colleges of North-East Nigeria. *International Journal of Innovative Scientific & Engineering Technologies Research, 8*(4), 30–40.

Hassan, S., Deba, A. A., & Saulawa, B. U. (2022). Development and standardization of an instrument for assessing carpentry and joinery students' practical skills in Nigerian senior secondary schools. *Canadian Journal of Educational and Social Studies, 2*(1), 73–85. https://doi.org/10.53103/cjess.v2i1.19

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing, 18*(3), 66–67. https://doi.org/10.1136/eb-2015-102129

Madadizadeh, F., & Bahariniya, S. (2025). Tutorial on internal consistency assessment by Cronbach's alpha and McDonald's omega. *Perioperative Care and Operating Room Management, 41*, 100568. https://doi.org/10.1016/j.pcorm.2025.100568

Mbaga, E. V., Yusuf, M. A., & Asukwo, A. E. (2024). Development and validation of a psychomotor skills assessment tool for electrical installation and maintenance works trade of technical colleges in Nigeria. *International Journal of Vocational Studies and Library Science, 2*(2), 1–21.

Nahadi, H., Firman, H., & Yulina, E. (2016). Performance assessment instrument to assess the senior high students' psychomotor for salt hydrolysis material. *AIP Conference Proceedings, 1708*, 040005. https://doi.org/10.1063/1.4941150

Nigerian Educational Research and Development Council. (2012). *Senior secondary school physics curriculum*. NERDC Press.

Nworgu, B. G. (2015). *Educational measurement and evaluation: Theory and practice*. University Trust Publishers.

Onunkwo, G. I. N. (2015). *Fundamentals of educational measurement and evaluation*. Cape Publishers International.

Prades, A., & Espinar, S. R. (2020). Laboratory assessment in chemistry: An analysis of the adequacy of the assessment process. *Assessment & Evaluation in Higher Education, 35*(4), 449–461. https://doi.org/10.1080/02602930902862857

Salim, K. R., Puteh, M., & Daud, S. M. (2022). Assessing students' practical skills in basic electronics laboratory based on psychomotor domain model. *Procedia - Social and Behavioral Sciences, 56*, 546–555. https://doi.org/10.1016/j.sbspro.2012.09.687

Simpson, E. J. (1972). *The classification of educational objectives in the psychomotor domain*. Gryphon House.

Yusuf, A., & Afolabi, A. O. (2018). Development and validation of an instrument for assessing students' laboratory skills in physics. *Journal of Educational Measurement and Evaluation in Nigeria, 9*(1), 76–85.