

**Reliability of alternative tests for evaluating strength and stamina
of muscles of the pectoral girdle (shoulder girdle)
of the military personnel**

Original scientific article
Strength and Endurance Development Tests

Alexander Piddubny

poddubnyag@gmail.com

<https://orcid.org/0000-0003-0957-2788>

University of Kharkiv Air Force Called Ivan Kozhedub,
61023, Kharkov-23, Sums kaya Street 77/79.

Capt. Federico Anibal Martinez Velez

<https://orcid.org/0000-0003-3315-2118>

mfederico_anibalv@hotmail.com

Comando de educación y Doctrina Militar terrestre del Ejército Ecuatoriano,
Jefe de la sección de investigación y experimentación de proyectos doctrinarios,
Universidad Nacional Mayor de San Marcos - Lima, Peru

Michael Anthony Martinez Velez

Zhanna Tzymbaliyk

<https://orcid.org/0000-0002-9129-5689>

zhanna.tzymbaliuk@gmail.com

University of Kharkiv Air Force Called Ivan Kozhedub,
61023, Kharkov-23, Sums kaya Street 77/79.

Sergey Palevich

<https://orcid.org/0000-0002-8304-1857>

junpolpsv@gmail.com

University of Kharkiv Air Force Called Ivan Kozhedub,
61023, Kharkov-23, Sums kaya Street 77/79.

Alexander Tkachuk

<https://orcid.org/0000-0002-0790-4104>

alextkachuk@i.ua

University of Kharkiv Air Force Called Ivan Kozhedub,
61023, Kharkov-23, Sums kaya Street 77/79.

Maxim Yarovy

<https://orcid.org/0000-0002-3338-1140>

maren1@ukr.net

University of Kharkiv Air Force Called Ivan Kozhedub,
61023, Kharkov-23, Sums kaya Street 77/79.

Artem Grishko

<https://orcid.org/0000-0001-9069-31443>

artem1641451@gmail.com

University of Kharkiv Air Force Called Ivan Kozhedub,
61023, Kharkov-23, Sumskaya Street 77/79.

Kharkiv National Air Force University
"Chief Marshal of Aviation Ivan Kozhedub",
Department of Physical Education,
Special Physical Training and Sports,
Kharkiv, Ukraine.
Sumskaya str. 77/79, Kharkiv, 61023, Ukraine

ABSTRACT

Purpose: This study is to evaluate the level of reliability in traditional and alternative tests, to evaluate the strength and endurance of muscles in the pectoral girdle.

Material and methods. The respondents study were 47 cadets aged 18-25. The reliability was established by two-fold evaluation by three randomly selected supervisors from among the professors of the departments for the procedure of PhF testing according to the requirements of the TPhEI and ACFT.

Results. Reliability was calculated by analyzing ANOVA, ICC and average differences using Bland-Altman methods. The results of the ANOVA don't confirm significant differences in the statistically results of supervisor's evaluation. The average difference between the 1st and 2nd measurements are -0.234 (± 1.96 SD = 3.887 – 4.35), 3.43 (20.47 – 13.62), 0.43 (0.95 – 0.87) and 0.19 (1.82 – 1.44) times for each of the 4 tests, respectively. The ICC showed the superior reliability of each controller for first, second measurements and the overall reliability of the ICC 1 (95% CI 0.999-1), 0.999 (0.999-1) alternative tests. The measurement error for PU, PshU, LT and Hand release pushups was 0.301, 1.27, 0.121 and 0.068 respectively. The SDC values were 0.85, 3.52, 0.35 and 0.188 for PU, PshU, LT and Hand pushups.

Conclusions. These tests demonstrate a practical, effective method of measuring the functional power.

Key words: reliability; tester; push up test; pull up; hand release pushups; leg tuck

Fiabilidad de pruebas alternativas para evaluar la fuerza y la resistencia de los músculos de la cintura pectoral (cintura escapular) del personal militar

RESUMEN

Propósito: Este estudio es evaluar el nivel de confiabilidad en las pruebas tradicionales y alternativas, para evaluar la fuerza y resistencia de los músculos de la cintura escapular.

Material y métodos. Los encuestados del estudio fueron 47 cadetes de entre 18 y 25 años. La confiabilidad fue establecida por evaluación doble por tres supervisores seleccionados al azar entre los profesores de los departamentos para el procedimiento de prueba PhF de acuerdo con los requisitos de TPhEI y ACFT.

Resultados. La confiabilidad se calculó analizando ANOVA, ICC y diferencias de promedio utilizando los métodos de Bland-Altman. Los resultados del ANOVA no confirman diferencias significativas en los resultados estadísticos de la evaluación del supervisor. La diferencia media entre la 1.^a y la 2.^a medición es -0,234 ($\pm 1,96$ SD = 3,887 – 4,35), 3,43 (20,47 – 13,62), 0,43 (0,95 – 0,87) y 0,19 (1,82 – 1,44) veces para cada una de las 4 pruebas, respectivamente. El ICC mostró la confiabilidad superior de cada controlador para la primera, segunda medición y la confiabilidad general del ICC 1 (IC 95% 0.999-1), 0.999 (0.999-1) pruebas alternativas. El error de medición para PU, PshU, LT y Hand release pushups fue 0,301, 1,27, 0,121 y 0,068 respectivamente. Los valores SDC fueron 0.85, 3.52, 0.35 y 0.188 para PU, PshU, LT y Hand pushups.

Conclusiones. Estas pruebas demuestran un método práctico y eficaz para medir la potencia funcional.

Palabras clave: fiabilidad; probador; prueba de flexiones; flexiones hacia arriba; flexiones con liberación manual; flexión de piernas

Artículo recibido: 03 marzo 2022

Aceptado para publicación: 20 marzo 2022

Correspondencia: poddubnyag@gmail.com

Conflictos de Interés: Ninguna que declarar

INTRODUCTION

Military conflict experience shows that in modern wars there is no «front line». Every soldier may find himself or herself in a position where he or she needs to shoot, move, overcome obstacles, lift, duck, carry cargo, and assist the wounded (Carlson, & Jaenen, 2012). In order to carry out these actions effectively, the military requires some basic level of sufficient muscle strength, dexterity, coordination, and endurance (Billing, Silk, Tofari, & Hunt, 2015; Batchelor, 2019). Muscle strength and stamina are important components of PhF (McManis, Baumgartner, & Wuest, 2000; Aandstad, 2020).

The necessary level of training is achieved by a planned, scientifically sound and systematic process for their physical improvement. The management of this process involves the time, and objective provision of information on the PhF of the military personnel. This task is solved by the system of verification and evaluation of PhF (TPhEI, 2014). It is based on assessment technology, tests and performance standards (Bompa, & Haff, 2009; Armstrong, Sinden, Sendsen, MacPhee, & Fischer, 2019).

Traditionally, the Armed Forces of Ukraine use the following exercise to evaluate the development of the shoulder girdle: Pull up (PU), Push up (PshU) (Temporary Instruction for Physical Training (TIPT)). Current research has shown that these tests have serious shortcomings in determining the readiness for military professional activities. This is how they evaluate the strength of the shoulder girdle, while combat missions require strength on the lower limbs.

Alternative tests are currently being tested in the armies of the world's leading nations, which evaluate the muscle strength and stamina of the upper/lower body parts (Peterson, 2015; Palevich, Poddubny, Tkachuk, & Zolochevsky, 2018). At the same time, the following exercises are used: bending and extending the arms in a lying down position, lifting the legs up to the elbows on the bar.

The results of these tests are evaluated by the examiners. The McManis et al. (2000) study identifies some of the problems that are common in the pull-up test, such as when participants take the SP or the «chin above the bar» position, which makes the assessment difficult. Moreover, it isn't possible to distinguish the physical development level in all the military personnel of the Armed Forces of Ukraine (Baumgartner, & Sharon, 2005).

Similarly, in the Lie-Down Bending and Extension Test, many experts are unable to determine the movement of the breast through the plane of right angles between the arms and the forearms, fully extending the arms, which leads to a low evaluation of the test's reliability (Baumgartner, Chung, & Hales, 2005). Therefore, objective information on the PhF of soldiers can only be obtained if the measurements can be performed reliably.

The study by Putranta and Supahar (2019) shows that in studying the overall score obtained from measurements between experts and the results of the evaluator's agreement, scores are almost always not identical. The reliability of the test is determined by the dispersion analysis of the expert estimates and the evaluation of the various components of dispersion (Arifin, Retnawati, & Putranta, 2020). Different indices for measuring agreement among several experts on the existence or absence of different measurement results can be interpreted as a correlation coefficient within a class (Rae, 1984).

Following current theoretical approaches in reliability studies, absolute and relative reliability measures need to be established (Hopkins, 2000; Weir, 2005). Relative reliability can be assessed by quantifying the correlation between repeated measurements, usually by obtaining an ICC (Shrout, & Fleiss, 1979).

Absolute reliability refers to the variability of estimates from test to test and is independent of the sample, as the range of individual estimates isn't taken into account. A common estimate of absolute variability is a standard measurement error (SEM), which is a measure within the subjective variation considered to be a «random variation in measurement when a person is checked repeatedly (Shrout, & Fleiss, 1979; Hopkins, 2000; Weir, 2005). Additional statistics called the Least Detectable Change (SDC) are increasingly used as a checkpoint for interpreting changes in scores. SDC indicates the smallest change in the estimate that occurs due to a real change in the estimate, not due to a measurement error.

This approach has already been used in psychometric research (Smits-Engelsman, & Niemeijer, 2011; Holm, & Tvetter, 2013; Proenca, Salomao, et al., 2014; Klijn, & Legemaat, 2015; Serbetar, & Loftesnes, 2019)

The objective of this study is to compare the reliability of traditional and alternative tests to evaluate the strength and endurance of the muscles of the upper body. Research materials can be used for training and testing guidance in the future.

METHODS

Sample

Respondents to this study were 47 male cadets aged 18-25 from Kharkiv National Air Force University "Chief Marshal of Aviation Ivan Kozhedub", The reliability of the tests was established by a double evaluation 8-10 days after the 1st check by three randomly selected supervisors from the PhE, ST and S Departments. All cadets were evaluated individually for the procedure of physical preparation verification in accordance with the requirements of the TIPT and the Army combat fitness test (ACFT).

The study was pre-approved. Each participant voluntarily provided a written informed consent prior to participation.

Data collection measuring instrument

The participants performed the exercise alternately, in that order by the controller, but before conducting the next test the subject had sufficient rest time.

Each supervisor received instructions on the conditions of exercise.

According to the terms of the TIPT, the PU on the bar is performed from the starting position (SP): hanging by the top of a bar with the arms straight, head straight, legs together. Bending the arms, raising the body in one move to the «chin above the bar» position. Go down in the SP without rocking. The controller only declared the score after fixing the SP of at least 1 and this was the permission to continue the exercise. They were not allowed to take their feet back to the SP, to perform the swerving movement of their body and legs, and to bend their knees. A slight slow deviation of the straight legs forward and the body from the stationary position was permitted.

The bending and extension of the arms in the lie-down support is performed from the SP: the lie-down stop, the arms parallel, the body straight, the legs together, relying on the socks. Bending the arms, lowering the straight body into the position where the breasts pass the plane of the right angle between the shoulders and the forearms, fully extending the arms to exit the SP, the invoice is declared after fixing the SP. During the bending and extension of the arms in the lie-down stop, it was permitted to stop for rest

in the SP. It was forbidden to bend and bend the body, to touch the floor with any part of the body, to tighten the legs. When the floor was touched simultaneously by the chest, the stomach, and the feet, the exercise stopped.

According to the ACFT conditions, the cadets make traditional bends and arm extensions in the lower position, but when in the lower position, release the hands from contact with the ground and then return to the previous position to perform another PshU. This allows the use of additional muscles in the shoulder girdle. Run time is 1 minute.

Like PU, cadets lift their feet up and down to touch their elbows as much as they can. This exercise strengthens the main muscles as it doubles the amount of force required compared to the traditional PU. Three experts independently recorded the results of each test.

Data analysis

Statistical analysis of the results was carried out using STATISTICA 10.0. and SPSS Statistics 17.0. The normal distribution was evaluated using the Shapiro-Wilk method. For the entire sample, the parameters of the descriptive statistics were calculated. Parametric indicators are presented as $M \pm SD$, where M (Mean) is the average, SD is the standard deviation. The reliability of the tests was determined on the basis of the recommendations described in the introduction (Hopkins, 2000). A variance analysis of expert estimates was carried out, and ICC were calculated. For the ICC calculation we have taken the form ICC, described in Shrout and Fleiss (1979) as ICC or bilateral with a random effect for absolute consent.

The consistency between the controllers was calculated using the ICC each for the first and second tests and all three controllers (overall reliability).

The Bland-Altman graphics were built to visually control and eliminate the presence of heterosecasticity (Bland, & Altman, 1986). Based on SEM, the smallest observed change (SDC) was calculated, which is the minimum difference that can be considered a real change between measurements with 95% confidence. SDC was calculated as $1,96 * SEM * \sqrt{2}$.

The calculation of the percentage error was as follows: $1,96 * \text{standard deviation of the mean difference between 1}^{\text{st}}$ and 2^{d} test / mean value of the test from the two measurements * 100%.

For all tests, a value level of $p \leq 0.05$ was used.

All reliability parameters were calculated using raw estimates because we suggested that it was more appropriate to obtain SEM and SDC values in real units of measure than in standard estimates.

RESULTS

As a result of a descriptive analysis of the data, it was obtained that the three controllers have different scores based on PU, PshU, Hand release pushups, LT. The results of the evaluation are presented as an average and a standard deviation in Table 1.

The highest average score with an average of 13.15 ± 6.91 times and 12.89 ± 5.86 times in the PU test on the 1st and 2^d test was given by the 2^d supervisor. The lowest score, averaging 10.91 ± 6.27 times in the first test, was obtained from the first supervisor and 11.98 times from the first and third supervisor. The results of the PU evaluation also show almost identical results, namely, the three supervisors have the results of the tests for which the highest average value is obtained from the third supervisor's evaluation with an average value of 37.53 ± 18.83 in the first test, and the lowest score with an average of 30.26 ± 13.56 received from the third supervisor in the second test. Hand release pushups and LT average scores are equal. These data show that there are differences between PU, PshU. However the ANOVA analysis presented in Table 2 doesn't confirm statistically significant differences in the results of the supervisor's evaluation. The calculated value F doesn't exceed the critical value F of $p > 0.05$.

The total ICC of the sample, based on the real units of measurement for each exercise is presented in Table 3. Table 3 shows that in four types of tests there are different correlation coefficients between supervisors. When evaluating the PU test, the reliability of each of the three experts was between good and excellent (the ICC lowest in the third controller was 0.942 and the highest 0.963 in the 2^d), while the overall reliability was excellent, ICC = 0.986 (0.978 0.992). The reliability of each expert in evaluating the PU test was between average and good (the ICC lowest in the 3^d controller was 0,838 and the highest 0.92 in the 1st), while overall reliability was excellent, ICC = 0.968 (0.95 0.981).

The reliability of each expert in evaluating the Hand release pushups and LT test is rated as excellent. ICC for all controllers is 0.999 and 1 with a correlation range of 0.999-1.

Based on Utkin's assumption (1978), ICC values below 0.699 to 0.600, 0.8 to 0.899, 0.9 to 0.949 and above 0.950 indicate low, acceptable, average, good and excellent reliability, respectively.

Given that ICC is highly dependent on sampling, we also calculated two sample-independent measures: SEM and least SDC. Since we used raw estimates, SEM and SDC are expressed in units of reference. To make SEM values comparable between exercises, they were also expressed as a percentage of the average (SEM%). As shown in Table 4, PU, PshU the largest SEM%. Accordingly, the SDC measures were also higher in these exercises.

The graph analysis of the results of the four measurement exercises by the three supervisors is shown in figures 1.

The Bland-Altman method for estimating the average difference between the 1st, 2^d measurements resulted in -0.234 times (± 1.96 SD = 3.887 – 4.35), 3.43 (20.47 – 13.62), 0.043 (0.95 – 0.87) and 0.19 (1.82 – 1.44) for each of the four tests, respectively. When estimating the graphics obtained by the Bland-Altman method, it is found that only three indicators (6.38%) went beyond 1.96 SD for PU, one indicator (2.13%) for PshU and three indicators (6.38%) for PshU for LT and six indicators (12.77%) for Hand release pushups. This suggests good reproducibility of all four tests.

It is thus clear from the analysis that the reproducibility of the new tests is very high. Therefore, new tests may be recommended for use in the assessment of the strength and stamina of the muscles of the upper shoulder girdle of servicemen.

DISCUSSION

Muscle strength and stamina are important components of the PhT of servicemen required for effective combat. According to D'Isanto et al. (2019), test scores are used to determine anthropometric and psychomotor profiles of a person, which are used to determine the goals needed to develop a learning program.

In this study, three different statistical analysis methods were used to assess the robustness of measurements by four tests performed for each observer and between observers in determining the level of strength and endurance of muscles in the upper shoulder girdle of servicemen.

The most preferred way to assess reliability is to use dispersion analysis followed by computation of ICC and Bland-Altman graphics for visual examination.

The results of the evaluation of several supervisors, including in one test, may differ. In order to reach a perfect agreement between the controllers it is necessary to carry out selection, their training to evaluate (Barnett, Beurden, Morgan, Lincoln, Zask, & Beard, 2009). The analysis based on the variances of the above results leads to the conclusion that there are no statistically significant differences in the average rating made by the three supervisors, indicating a high degree of preparedness and consistency among them.

The findings are consistent with those of Negrete et al. (2010), which conducted studies on men and women, and Benny & Matthew (2001), which conducted studies on schoolchildren.

The results are slightly higher than those of Arifin, Retnawati & Putranta (2020), which investigated the Indonesian Air Force. The explanation for this is that the cadets have more experience than the soldiers in the performance of these exercises (they participate more in monitoring activities: once a month control of the plan of sports mass work, during scheduled classes, at the end of the semester exam or credits). The description of performing exercises in TPhEI is more detailed than As many as five randomly selected testers came from the Air Force Physical Development unit and were experienced and often involved in PhF testing. The requirements for performing alternative tests have fewer problematic movements and positions that the controller assesses.

The test reliability of the PU, PshU, Hand release pushups and LT tests was excellent. The ICC showed the superior reliability of each controller for 1st, 2^d measurements and the overall reliability of the ICC 1 (95% CI 0.999-1), 0.999 (0.999-1) alternative tests.

However, in terms of psychometric theory, both high and low ICC values should be taken with caution because, as stated in Weir (2005), «large ICC can mask poor consistency between tests when variability between subjects is high» and «and conversely, a low ICC can be detected even if the variability from test to test is low if the variability between subjects is low». Weir (2005) also pointed out the importance of the source of the error, which should be briefly addressed. Namely, the term error in ANOVA expresses the interaction of subjects and tests, where a small error may reflect that estimates change in a similar way in repeated tests, which may lead to a significant test effect, which means there's some kind of systematic error. On the contrary, a random error may exist in the data when changes between tests are not harmonized.

Since the two-sided model allows split error, we checked the average squares in the ANOVA output without finding the effect for the overall evaluation of the tests.

The average difference between the 1st, 2^d measurements is - 0.234 times (± 1.96 SD = 3.887 – 4.35), 3.43 (20.47 – 13.62), 0.043 (0.95 – 0.87) and 0.19 (1.82 – 1.44) for each of the four tests, respectively.

SEM and SDC were calculated. SEM (as an estimate of absolute reliability) indicates an expected error in measuring an individual estimate, expressed in real units of measure, while SDC represents a confidence interval around the error. The measurement error for PU, PshU was 0.301 and 1.27. For LT and Hand release pushups 0.121 and 0.068 respectively. The SDC values were 0.85 and 3.52 for PU, PshU, and for LT and Hand pushups 0.35 and 0.188. The minimum detectable change for PU, PshU was slightly higher than Leg tuck and Hand release pushups. The SDC values of traditional tests were also higher, resulting in a larger measurement error.

This study confirmed the findings of the scientists (Baumgartner, & Gaunt, 2005; Barnett et al., 2009). Arifin, Retnawati & Putranta (2020) that, in order to improve test reliability, it is necessary to clearly describe the problem movements, define the position so that the controller can accurately evaluate the exercise. Before testing, train and instruct controllers to use auxiliary devices to facilitate measurement, apply alternative forms of testing.

CONCLUSION

The results of the applied statistical methods (analysis of ANOVA, ICC, and average differences (Bland-Altman method)) have shown the excellent reliability of alternative tests of assessment of strength level and endurance of muscles of the upper shoulder girdle of servicemen. These tests, like traditional ones, demonstrate a practical and effective method of measuring the functional power of the upper shoulder girdle. However, since only one of the traditional tests examined is used in testing and the alternative tests are used both, their joint application increases the reliability of the estimates.

Conflict of Interest

The authors declare that there is no conflict of interest.

REFERENCES

A

a

- Arifin, S., Retnawati, H., & Putranta, H. (2020). Indonesian air force physical tester reliability in assessing one-minute push-up, pull-up, and sit-up tests. *Sport Mont*, 18 (2), 89-93. doi: 10.26773/smj.200614
- Barnett L., Beurden, E., Morgan, P. J., Lincoln, D., Zask, A., & Beard, J. (2009). Association Between Field-based in Muscular Fitness Field Tests and Assessment. *Research Quarterly for Exercise and Sport*, 80 (2), 363-368. <https://doi.org/10.1080/02701367.2009.10599571>
- Batchelor, J. (2019). *Applicability of the army physical fitness test in the contemporary operating environment*. Kansas.
- Baumgartner, Ted A., & Gaunt, Sharon j. (2005). Construct Related Validity for the Baumgartner Modified Pull-Up Test. *Measurement in Physical Education and Exercise Science*, 9, 1, 51-60.
- Billing, D., Silk, A., Tofari, P., & Hunt, A. (2015). Effects of Military Load Carriage on Susceptibility to Enemy Fire During Tactical Combat Movements. *J. Strength Cond Res.*, 29 (11), 134-8.
- Bisca, G. W., Proenca, M., Salomao, A, & al. (2014). Minimal detectable change of the London Chest Activity of Daily Living Scale in patients with COPD. *J. Cardiopulm Rehabil Prev*, 34, 213-216. [PubMed] [Google Scholar]
- Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1 (8476), 307-10. [PubMed] [Google Scholar]
- Bodden, Andrew, & Baghurst, Timothy (2013). A practical method for determining muscular strength in adolescents over time. *Oklahoma AHPERD Journal*, 50, 75-94.
- Bompaş T. O., & Haff, G. G. (2009). *Periodization: Theory and methodology of training. 5-th Edition*. Champaign, IL, USA: Human Kinetics.
- Carlson, M., & Jaenen, S. (2012). The development of a preselection physical fitness training program for Canadian special operations regiment applicants. *J. Strength Cond. Res.*, 26, 2-14. [Google Scholar] [CrossRef] [PubMed]

S

o

i

- Cogley, R. M., Archambault, T. A., Fibeger, J. F., Koverman, M. M., Youdas, J. W., & Hollman, J. H. (2005). Comparison of muscle activation using various hand positions during the push-up exercise. *Journal of Strength and Conditioning Research*, 19 (3), 628-633.
- D'Isanto, T., D'Elia, F., Raiola, G., & Altavilla, G. (2019). Assessment of sports performance: Theoretical aspects and practical indications. *Sport Mont*, 17 (1), 79-82. <https://doi.org/10.26773/smj.190214>
- Daniel P., Armstrong, Kathryn E., Sinden, Jonathan, Sendsen, Renée, S., MacPhee, & Steven L., Fischer (2019). The Ottawa Paramedic Physical Ability Test: test-retest reliability and analysis of sex-based performance differences. *Ergonomics*, 62, 8, 1033-1042. DOI: [10.1080/00140139.2019.1618501](https://doi.org/10.1080/00140139.2019.1618501)
- Holm, I., Tveter, A. T., Aulie, V. S., & Stuge, B. (2013). High intra- and inter-rater chance variation of the movement assessment battery for children 2, ageband 2. *Res. Dev. Disabil*, 34, 795-800. [CrossRef] [PubMed]
- Hopkins, W. G. (2000). Measures of Reliability in Sports Medicine and Science. *Sports Med.*, 30, 1-15. [CrossRef] [PubMed]
- Klijn, P., Legemaat, M., Beelen, A., Keimpema, A. V., Garrod, R., Bergsma, M., Paterson, B., Stuijzand, A., & van Stel, H. (2015). Validity, Reliability, and Responsiveness of the Dutch Version of the London Chest Activity of Daily Living Scale in Patients With Severe COPD. *Medicine*, 94 (49), 2191. <https://doi.org/10.1097/MD.0000000000002191>
- McManis, B. G., Baumgartner, T. A., & Wuest, D. A. (2000). Objectivity and reliability of the 90° push-up test. *Measurement in Physical Education and Exercise Science*, 4 (1), 57-67.
- Negrete, R. J., Hanney, W. J., Kolber, M. J., Davies, G. J., Ansley, M. K., McBride, A. B, & Overstreet, A. L. (2010). Reliability, minimal detectable change, and normative values for tests of upper extremity function and power. *J. Strength Cond Res.*, 24 (12), 3318-25. doi: 10.1519/JSC.0b013e3181e7259c. PMID: 21088548
- Peterson, David, Msc. (2015). Modernizing the Navy's Physical Readiness Test: Introducing the Navy General Fitness Test and Navy Operational Fitness Test. *The Sport Journal*.

- Provisional Guidelines for Physical Training in the Armed Forces of Ukraine (TNFP-2014). Ky`yiv, 160. In Ukrainian.
- Putranta, H., & Supahar, S. (2019). Development of physics-tier tests (PysTT) to measure students' conceptual understanding and creative thinking skills: A qualitative synthesis. *Journal for the Education of Gifted Young Scientists*, 7 (3), 747-775. <https://doi.org/10.17478/jegys.587203>
- Rae, G. (1984). On measuring agreement among several judges on the presence or absence of a trait. *Educational and Psychological Measurement*, 4, 247-253.
- Romain, Benny, & Mahar, Matthew (2001). Norm-Referenced and Criterion-Referenced Reliability of the Push-Up and Modified Pull-Up. *Measurement in Physical Education and Exercise Science*, 5. 67-80. DOI: 10.1207/S15327841MPEE0502_1
- S., Palevich, A., Poddubny, A., Tkachuk, & V., Zolochevsky (2018). State of problems and directions for improvement special physical training serviceman of the air Force of the Armed Forces Of Ukraine. *Sport science of Ukraine*, 1 (83), 15-25.
- Serbetar, Ivan, Loftesnes, Jan M., & Mamen, Asgeir (2019). Reliability and Structural Validity of the Movement Assessment Battery for Children-2 in Croatian Preschool Children. *Sports (Basel, Switzerland)*, 7. 10.3390/sports7120248.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychol. Bull*, 86, 420-428. [CrossRef] [PubMed]
- Ted A., Baumgartner, Suhak, Oh, Hyuk, Chung, & Derek, Hales (2002). Objectivity, Reliability, and Validity for a Revised Push-Up Test Protocol. *Measurement in Physical Education and Exercise Science*, 6, 4, 225-242. DOI: [10.1207/S15327841MPEE0604_2](https://doi.org/10.1207/S15327841MPEE0604_2)
- Utkin, V. L. (1978). *Measurements in sports*. Moskow, GTSOLIFK. 200.
- Weir, J. P. (2005). Quantifying Test-Retest Reliability Using the Intraclass Correlation Coefficient and the SEM. *J. Strength Cond. Res.*, 19, 231. [CrossRef] [PubMed]

ANEXOS

Table 1. Results of Pull Up, Push Up, Hand release pushups, Leg tuck Tests

Test	Testing 1 (Mean ± SD)			Testing 2 (Mean ± SD)		
	Tester 1	Tester 2	Tester 3	Tester 1	Tester 2	Tester 3
Pull up	10.91 ± 6.27	13.15 ± 6.91	12.09 ± 6.80	11.98 ± 6.03	12.89 ± 5.86	11.98 ± 5.89
Push up	34.98 ± 16.26	36.6 ± 19.06	37.53 ± 18.83	34.45 ± 12.62	34.13 ± 13.74	30.26 ± 13.56
Hand release pushups	32.85 ± 15.12	32.94 ± 15.1	32.66 ± 15.13	32.64 ± 14.66	32.57 ± 14.52	32.66 ± 14.64
Leg tuck	12.4 ± 7.4	12.45 ± 7.4	12.49 ± 7.34	12.4 ± 7.27	12.4 ± 7.24	12.4 ± 7.23

Table 2. ANOVA Analysis Results for Differences in Assessment

Test	F critical	F hit	Sig
Pull up	2.25	0.742	0.593
Push up		1.185	0.317
Hand release pushups		0.042	1
Leg tuck		0.001	1

Table 3. Intraclass correlation coefficients (ICC)

Test	ICC (± 95% CI)			
	Tester 1	Tester 2	Tester 3	Tester 1-3
Pull up	0.959 (0.910 0.980)	0.963 (0.935 0.980)	0.942 (0.895 0.967)	0.986 (0.978 0.992)
Push up	0.92 (0.857 0.956)	0.873 (0.773 0.929)	0.838 (0.55 0.928)	0.968 (0.95 0.981)
Hand release pushups	0.999 (0.998 0.999)	0.999 (0.998 0.999)	0.998 (0.997 0.999)	1 (0.999 1)
Leg tuck	0.999 (0.998 0.999)	0.998 (0.996 0.999)	0.998 (0.996 0.999)	0.999 (0.999 1)

Table 4. Standard error of measurement (SEM) and smallest detectable change (SDC) values for each age

Statistics options	Pull up	Push up	Hand release pushups	Leg tuck
SEM	0.301	1.27	0.121	0.068
SEM%	33.87	49.18	4.97	7.32
SDC95	0.85	3.52	0.35	0.188

Figure 1. Graphical data showing the differences (Y-axis) between control 1 and control 2 (separated by at least 1 week) compared to the mean (X-axis) of the same two dimensions according to the Bland-Altman method (1986). The graphics include: a – Pull up; b – Push up. c – Leg tuck and d – Hand release pushups when re-measured by three controllers

