

EXPERIMENTAL VERIFICATION OF THE MODEL FOR DEVELOPING FOREIGN LANGUAGE WRITING COMPETENCE IN FUTURE MASTERS IN NATIONAL SECURITY

Viacheslav Lobko ^A; Svitlana Mamchur ^B

^A e-mail: 37hogan@gmail.com, ORCID ID: <https://orcid.org/0009-0002-0907-0083>

^B e-mail: svitlana.svetljachok@gmail.com, ORCID ID: <https://orcid.org/0009-0004-6365-1799>

Yevheniy Berezhnyak Military Academy, Kyiv, Ukraine

Received: June 10, 2026 | **Revised:** June 24, 2026 | **Accepted:** June 30, 2026

UDC 378.147:81'243:355.231.2

DOI: <https://doi.org/10.33445/psssj.2026.7.2.4>

Abstract

This article reports the design, procedure and results of a pedagogical experiment conducted to verify the effectiveness of the authors' model for developing foreign language writing competence in future masters in national security who study a foreign (Croatian) language from scratch. The study aims to answer two key questions: (1) how should the experimental verification of such a model be organised under the constraints of small language groups typical of higher military educational institutions (HMEIs)? (2) does instruction under the proposed model and pedagogical conditions produce a statistically significant advantage over traditional instruction? The model comprises target, methodological, content, procedural and diagnostic-resultative blocks and is implemented through a set of pedagogical conditions. A natural open experiment covered four stages – ascertaining, intermediate, formative and interpretative – with the formative stage running through the 2021–2023 and 2023–2025 academic years in two experimental (EG 1, 2) and two control (CG 1, 2) groups formed from successive trainee intakes. Learning outcomes were diagnosed through two STANAG 6001-based test administrations. Given the small samples, the non-parametric Mann-Whitney U test was applied, computed in SPSS and verified manually. The empirical value $U = 0$ for both administrations, with an exact two-tailed significance of $p = 0.029$ ($p < 0.05$), warranted rejection of the null hypothesis: the writing competence of EG trainees proved statistically significantly higher than that of their CG peers. The effectiveness of the pedagogical conditions was further corroborated by a group expert assessment (18 experts, group representativeness $K_{gr} = 0.72$), in which over 90 % of the experts confirmed the proposed conditions. The findings offer HMEIs an evidence-based, transferable framework for ab initio foreign language instruction aligned with NATO language policy, and demonstrate a statistically rigorous verification procedure suitable for the small cohorts in which military language training is typically delivered.

Key words: Foreign Language Writing Competence; Military Education; Pedagogical Experiment; Pedagogical Conditions; Mann-Whitney U Test; STANAG 6001.

Introduction

The full-scale armed aggression of the Russian Federation against Ukraine has made the profound transformation of the military education system, and its alignment with the standards, norms and practices of NATO member states, a matter of urgency (Cabinet of Ministers of Ukraine, 1997). The participation of the Armed Forces of Ukraine in multinational activities, the roll-out of the NATO Defence Education Enhancement Programme (NATO, n.d.) and the requirements of the Alliance's language policy, codified in the Standardized Language Profiles (SLP) under STANAG 6001 (NATO Standardization Office, n.d.), impose qualitatively new demands on the foreign language training of

the officer corps. For specialists in the field of national security, writing competence is of particular consequence: written communication – staff correspondence, dispatches, reports and analytical papers – is the dominant mode of professional interaction in the international security environment.

A specific challenge lies in developing this competence when master's-level trainees begin studying a foreign language from scratch (no proficiency level) and instructional time is severely constrained by intensive professional training. Meeting this challenge requires a scientifically substantiated model for developing foreign language writing competence, an optimised set of pedagogical conditions for its implementation and, crucially, evidence-based verification of its effectiveness through a pedagogical experiment with statistical substantiation appropriate for the small samples typical of language groups in HMEIs.

This study addresses that requirement. Its purpose is to describe the design and course of the experimental verification of the authors' model for developing foreign language writing competence in future masters in national security, and to carry out a statistical analysis and interpretation of the results of the formative experiment. Specifically, it poses two interconnected questions: (1) how should such verification be organised under the constraints of small, successive trainee intakes, and (2) does instruction under the proposed model and pedagogical conditions yield a statistically significant advantage over traditional instruction?

Literature Review

The essence, content, aims and methods of the pedagogical experiment have been examined by V. Yagupov (2019), O. Zhosan (2008), N. Tverezovska and V. Sydorenko (2013), O. Dubaseniuk (2016) and E. Panasenko (2015). The methodological foundations of pedagogical research are set out comprehensively in the works of S. Honcharenko (2010) and O. Klymeniuk (2006), while the requirements for designing and evaluating research in education are articulated in the seminal work of J. Fraenkel, N. Wallen and H. Hyun (2012). The toolkit for the mathematical-statistical processing of experimental data, in particular for small samples, has been characterised by S. Mitina (2024); the capabilities of the SPSS package for data analysis are described by A. Horbachyk and S. Salnikova (2008). The methodology of group expert assessment and the determination of expert competence have been substantiated by O. Velychko and co-authors (Velychko et al., 2015). The methodology and methods of dissertation research on competence development within the higher military education system were addressed in an earlier publication (K. Mamchur, 2024).

The normative dimension of military language education is defined by the Alliance's standardization framework: STANAG 6001, Edition 5, sets the agreed proficiency levels for curriculum and test development and for recording Standardized Language Profiles (NATO Standardization Office, n.d.), while the Common European Framework of Reference for Languages provides the action-oriented descriptors that inform civilian and military curricula alike (Council of Europe, 2020). At the same time, a holistic model for developing foreign language writing competence in future masters in national security who acquire the language from scratch – together with the results of its experimental verification and statistically substantiated evidence on small samples – has received only fragmentary coverage in the scholarly discourse and has remained outside researchers' attention. The present study seeks to fill this gap.

Materials and methods

The pedagogical experiment was designed to test, in practice, the effectiveness of the authors' model for developing foreign language writing competence in future military specialists. It was conducted under conditions ordinary for a higher education institution, in the course of regular language training and without any selection of participants – hence, following the classifications of V. Yagupov (2019) and O. Zhosan (2008), it qualifies as a natural open experiment. The experimental

study proceeded in four stages. At the ascertaining stage, the existing approaches to developing foreign language writing competence and the current state of that competence in HMEIs were established. The intermediate stage involved designing the model (substantiating the approaches, principles, methods, techniques and the system of exercises), optimising the pedagogical conditions and selecting teaching materials. The formative experiment, conducted during the 2021–2023 and 2023–2025 academic years, was aimed at trialling the experimental model and the pedagogical conditions for developing writing competence, and at diagnosing its development. The interpretative stage, carried out in 2025–2026, consisted in the statistical substantiation of the data obtained, the step-by-step testing of the research hypothesis and the formulation of conclusions.

The formative experiment involved two experimental (EG 1, EG 2) and two control (CG 1, CG 2) groups of four trainees each; the assignment of study groups reflected successive intakes of trainees enrolled to study Croatian. Since the trainees began the language at the no-proficiency level, there was no need for an ascertaining assessment to detect between-group differences in competence at the outset. Instead, to confirm the soundness of the group assignment, the trainees' academic performance in the third semester (immediately preceding the second foreign language course) was analysed: the difference in mean grades across disciplines averaged 0.2 points between CG 1 and EG 1 and 0.4 points between CG 2 and EG 2, which does not exceed 5 % of overall mean performance and attests to the comparability of the groups on this indicator.

Instruction under the experimental pedagogical conditions lasted four semesters and comprised three stages – basic initial, basic advanced and professional. The analysis of the formative experiment consisted in the statistical substantiation and interpretation of its results and in the step-by-step testing of the research hypothesis, namely that conducting practical classes in accordance with the experimental pedagogical conditions would foster the development of foreign language writing competence in master's-level trainees. To this end, specific hypotheses were advanced at each phase of the experiment, whose confirmation or refutation validated the logic of our reasoning. The null hypothesis (H_0) held that the writing competence of EG trainees (Y_1) after instruction under the experimental technology is not higher than that of CG trainees (Y_2), i.e. $H_0: Y_1$ is not higher than Y_2 ; the alternative hypothesis (H_1) held that the writing competence of EG trainees is higher than that of CG trainees, i.e. $H_1: Y_1$ is higher than Y_2 . As the experiment involved two experimental and two control groups, the test was performed twice.

The development of foreign language writing competence was evaluated against four criteria with corresponding indicators (Table 1), aligned with the STANAG 6001 writing descriptors (NATO Standardization Office, n.d.) and the Common European Framework of Reference for Languages (Council of Europe, 2020).

Table 1: Criteria and Indicators for Evaluating the Development of Foreign Language Writing Competence

Criterion	Key indicators	Instruments
Linguistic-knowledge	orthographic accuracy; range and precision of vocabulary (including military-professional terminology); grammatical accuracy of written speech	STANAG 6001-based test tasks
Textual-discursive	text organisation; coherence and logical flow; conformity to the genre conventions of the written document	analysis of written products
Professional-pragmatic	ability to accomplish professional writing tasks (staff correspondence, dispatches, reports); appropriateness of style and register to the professional communicative situation	quasi-professional writing tasks; expert assessment

Criterion	Key indicators	Instruments
Motivational-reflexive	stability of motivation to improve written speech; capacity for self-assessment and error-driven revision	surveys; language portfolio

Source: developed by the authors

Quantitative indicators were used to confirm the effectiveness of the pedagogical conditions: experimental verification was performed by comparing the assessment data of EG 1, 2 and CG 1, 2 using statistical methods of substantiation. In selecting the statistical criteria, we proceeded from the classification of research tasks and from the power of a criterion – its capacity to detect differences where they exist. The selection rested on the simplicity of the criterion, its broad range of application (including small samples) and its applicability to samples of unequal size. Statistical criteria were selected following the general classification of research tasks and the methods for solving them (Mitina, 2024), as shown in Table 2.

Table 2: Selection of Statistical Criteria for Hypothesis Testing

Task	Conditions	Methods
Detecting differences in the level of the attribute under study	two independent samples of subjects	Mann-Whitney U test; ϕ^* criterion (Fisher's angular transformation)
Detecting differences in the level of the attribute under study	sample does not exceed 8 subjects	Mann-Whitney U test

Source: compiled by the authors after Mitina (2024)

The conventional 5 % significance level was adopted. The results were analysed with the SPSS statistical package for the social sciences (Horbachyk & Salnikova, 2008), which supports all the operations required for the present data and allows the appropriate statistical criterion to be selected; the correctness of the results was additionally verified using mathematical formulas.

Results

The comparative analysis drew on STANAG 6001-based test diagnostics administered after the basic advanced stage (first assessment, maximum 63 points) and the professional stage (second assessment, maximum 99 points) of competence development. The test results are presented in Table 3 (personalised records have been aggregated; trainees are coded).

Table 3: Results of Test Diagnostics of Foreign Language Writing Competence in Future Military Specialists

Group	Trainee (code)	First assessment (max 63 points)	Second assessment (max 99 points)
CG 1	C1.1	57	87
	C1.2	54	83
	C1.3	56	88
	C1.4	57	88
	Mean score	56.00	86.50
CG 2	C2.1	55	86
	C2.2	58	90

Group	Trainee (code)	First assessment (max 63 points)	Second assessment (max 99 points)
	C2.3	54	81
	C2.4	57	88
	Mean score	56.00	86.25
EG 1	E1.1	62	96
	E1.2	60	93
	E1.3	61	95
	E1.4	59	92
	Mean score	60.50	94.00
EG 2	E2.1	59	93
	E2.2	61	94
	E2.3	60	94
	E2.4	61	96
	Mean score	60.25	94.25

Source: compiled by the authors from the assessment results

The test results confirmed the higher level of foreign language writing competence attained by the trainees of the experimental groups. The group means for both assessments are compared in Figure 2,

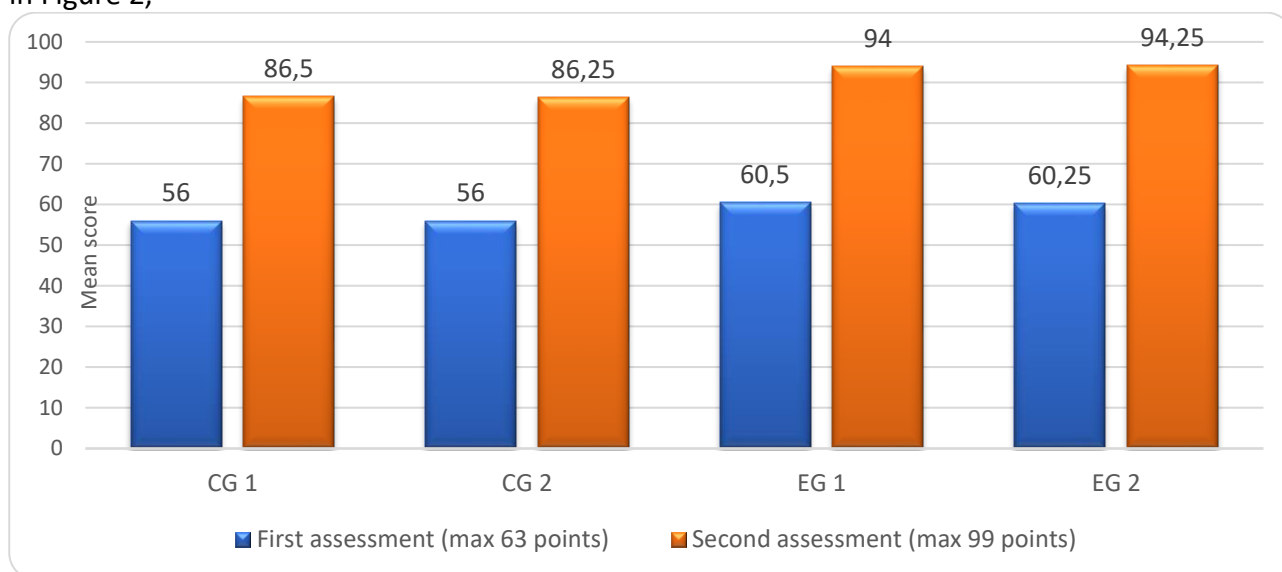


Figure 2: Comparison of Mean Scores of CG 1, 2 and EG 1, 2 Trainees by Assessment

and the dynamics of relative success (as a percentage of the maximum score of each assessment) in Figure 3.

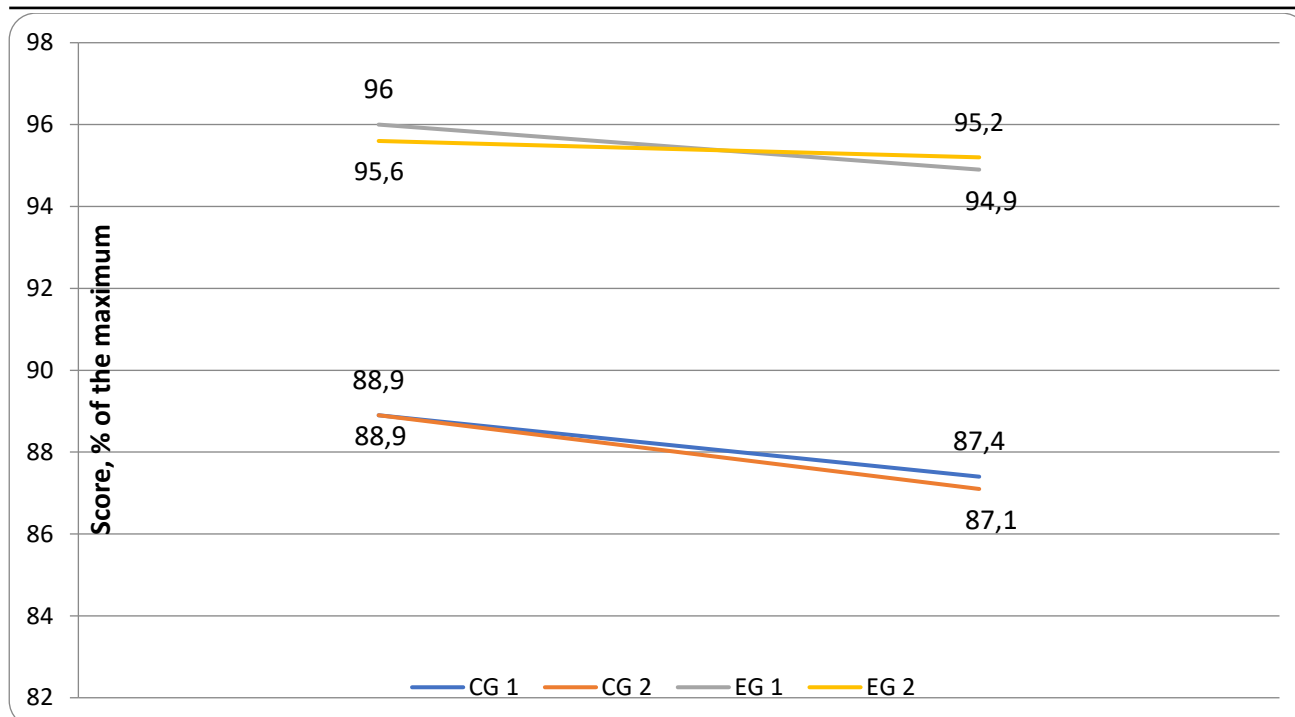


Figure 3: Dynamics of the Relative Success of CG 1, 2 and EG 1, 2 Trainees Between Assessments, % of the Maximum Score

As the data show, the mean results of EG 1 and EG 2 exceed those of CG 1 and CG 2 by 4.25–4.50 points at the first assessment and by 7.50–8.00 points at the second; in relative terms, the advantage of the experimental groups amounts to 7.1–7.5 percentage points, with no overlap between the ranges of individual results across the groups. The analysis of these indicators established that both the mean and the total ranks of the scores in EG 1, 2 are higher than in CG 1, 2 (Table 4).

Table 4: Ranks of the Test Diagnostics in the Comparison of the Experimental and Control Groups

Group	First assessment			Second assessment		
	number of trainees	mean rank	sum of ranks	number of trainees	mean rank	sum of ranks
CG 1	4	2.50	10.00	4	2.50	10.00
CG 2	4	2.50	10.00	4	2.50	10.00
EG 1	4	6.50	26.00	4	6.50	26.00
EG 2	4	6.50	26.00	4	6.50	26.00

Source: computed by the authors in SPSS

For the statistical substantiation of the diagnostic results, the composite Mann-Whitney U statistic was computed, providing grounds either to accept or to reject the null hypothesis. The results obtained with SPSS were verified using the formula:

$$U = n_1 \cdot n_2 + n_x \cdot (n_x + 1) / 2 - T_x, \quad (1)$$

where n_1 – is the number of trainees in the first group;
 n_2 – is the number of trainees in the second group;
 T_x – is the larger of the two rank sums;

n_x – is the number of trainees in the group with the larger rank sum.

The computed statistics are presented in Table 5.

Table 5: Computed Statistics of Significance Testing

Statistic	First assessment	Second assessment
Mann-Whitney U	0.000	0.000
Wilcoxon W	10.000	10.000
Asymptotic significance (2-tailed)	0.021	0.020
Exact significance (2-tailed)	0.029	0.029

Source: computed by the authors in SPSS

The critical values of the Mann-Whitney U statistic for the chosen significance level were determined from the table of critical values for the given n_1 and n_2 . If the obtained value of U is less than or equal to the tabulated value, a significant difference between the levels of the attribute in the samples under comparison is recognised (the alternative hypothesis is accepted); if the obtained value exceeds the tabulated one, the null hypothesis is retained. The tabulated U value for samples of four trainees in EG 1, 2 and four in CG 1, 2 at the significance level of $p = 0.05$ equals zero. Since $U_{emp} = 0 = U_{cr}$, with an exact two-tailed significance of $p = 0.029 < 0.05$, the hypothesis of a non-substantial difference between the diagnostic results of EG 1, 2 and CG 1, 2 trainees is rejected: the level of foreign language writing competence of EG 1, 2 trainees is statistically significantly higher than that of CG 1, 2 trainees.

Group expert assessment served as an additional method of verifying the effectiveness of the pedagogical conditions. The required number of experts was determined following the recommendations of Velychko et al. (2015) and the formula $N \geq j \cdot (d/\epsilon)^2$, where d is the range of individual assessments; ϵ is the specified error of the collective expert assessment; and j is a coefficient equal to 0.09 for a confidence probability of $0.8 < g < 0.9$. The calculations demonstrated that 17–18 experts suffice for decision-making at a confidence probability of at least $g = 0.8$, which accords with the view of the methodology's authors. The expert group comprised 18 members: 15 faculty members of HMEIs (including one Doctor of Political Science, five PhDs in Education, four PhDs in Political Science, two PhDs in Engineering, and one PhD each in Philology, Philosophy and Psychology) and three administrators closely involved in the educational process (PhDs in Engineering). Experts were selected by the documentary method of appraising their qualities against the indicators of competence, engagement, efficiency and objectivity.

Each expert's competence was determined as the ratio of the scores obtained to the maximum attainable across the questionnaire items ($K_i = \Sigma x / \Sigma x_{max}$), and the group's competence as the arithmetic mean of the individual indices ($K_{gr} = (1/n) \cdot \Sigma K_i$) (Velychko et al., 2015). The representativeness of the expert group amounted to $K_{gr} = 0.72$, attesting to the sufficiently high competence of the selected experts. The experts completed a 30-item multiple-choice questionnaire evaluating the effectiveness of the pedagogical conditions from a methodological standpoint; a single-round survey was conducted without direct interaction among the experts. Over 90 % of the respondents confirmed the effectiveness of the proposed pedagogical conditions.

Discussion

The findings confirm the research hypothesis: purposeful instruction under the authors' model and pedagogical conditions produces a consistent and statistically significant advantage in writing competence over traditional instruction. The advantage is observable at both assessment points, amounts to 7.1–7.5 percentage points in relative terms and, notably for samples of this size, is

accompanied by a complete absence of overlap between the individual score ranges of the experimental and control groups. The convergence of three independent lines of evidence – descriptive comparison, non-parametric inference and group expert assessment – strengthens the internal validity of the conclusion, in line with the requirement for methodological triangulation in experimental pedagogy (Fraenkel et al., 2012; Yagupov, 2019).

Two methodological choices deserve emphasis. First, the natural open design (Zhosan, 2008) preserved the ecological validity of the results: instruction proceeded under the ordinary conditions of an HMEI, with successive intakes rather than artificially constructed cohorts, so the observed effect reflects what the model achieves in real institutional practice. Second, the statistical apparatus was matched to the constraint that defines military language training – very small groups. The Mann-Whitney U test with exact significance is appropriate precisely where parametric assumptions cannot be defended (Mitina, 2024), and manual verification of the SPSS output (Horbachyk & Salnikova, 2008) excluded computational artefacts. Taken together, these choices offer a transferable verification procedure for other researchers working with the small cohorts typical of specialised military education.

The study also has limitations that bound the generalisability of its conclusions. The samples, although sufficient for exact non-parametric inference, are small (four trainees per group); the experiment covered a single language of instruction, a single institution and one type of speech activity (writing); and diagnostics relied on two assessment points. The rejection of the null hypothesis therefore warrants confidence in the direction and significance of the effect, but estimates of its magnitude should be treated with caution. These limitations chart the agenda for further work: replication on larger cohorts with parametric verification, extension of the model to other languages and skills, and longitudinal tracking of graduates' language performance in professional settings.

Conclusions

1. The experimental verification of the model and the pedagogical conditions for developing foreign language writing competence in future masters in national security was carried out within a natural open pedagogical experiment covering the ascertaining, intermediate and formative stages and the stage of result interpretation. The formative experiment was conducted during the 2021–2023 and 2023–2025 academic years in two experimental and two control groups formed from successive intakes of trainees studying Croatian from scratch; the comparability of the groups was confirmed by an analysis of academic performance in the preceding semester (the divergence in mean grades did not exceed 5 %).

2. Instruction under the experimental pedagogical conditions lasted four semesters and comprised three stages – basic initial, basic advanced and professional. Two test diagnostics administered with STANAG 6001-based tasks established that both the mean and the total ranks of the scores in EG 1, 2 exceed those in CG 1, 2, and that the EG mean scores surpass the CG indicators at both assessments, with no overlap between the ranges of individual results.

3. Given the small sample sizes, statistical substantiation was performed with the non-parametric Mann-Whitney U test, computed in SPSS and verified by mathematical formulas. Since $U_{\text{emp}} = 0$ against the tabulated value $U_{\text{cr}} = 0$ for samples of four and four subjects ($p = 0.05$), with an exact two-tailed significance of $p = 0.029$, the null hypothesis was rejected: the level of foreign language writing competence of the experimental groups' trainees is statistically significantly higher than that of the control groups.

4. The effectiveness of the pedagogical conditions was further confirmed by group expert assessment: the representativeness of the 18-member expert group amounted to $K_{\text{gr}} = 0.72$, and over 90 % of the experts confirmed the effectiveness of the proposed conditions. The analysis of the

empirical data, the expert assessment and the mathematical statistics jointly corroborate the effectiveness of introducing the authors' model into the professional training of future military specialists.

5. Prospects for further research lie in developing methodological guidance for implementing the experimentally validated pedagogical conditions in the educational process of HMEIs, in adapting the model to other types of speech activity and languages of instruction, and in verifying the findings on larger samples with parametric statistical criteria.

Funding

This study received no specific financial support.

Competing interests

The authors declare that they have no competing interests.

Use of Artificial Intelligence

Artificial intelligence was used for language editing, translation, and technical verification of the manuscript.

References

- Cabinet of Ministers of Ukraine. (1997). On the transformation of the military education system: Decree No. 1410 of December 15, 1997 (as amended on January 4, 2023). URL: <https://zakon.rada.gov.ua/laws/show/1410-97-п#Text> [in Ukrainian].
- Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment – Companion volume. Council of Europe Publishing. URL: <https://www.coe.int/en/web/common-european-framework-reference-languages>
- Dubaseniuk, O. (2016). Methodology and methods of scientific-pedagogical research: Study guide. Polissia. [in Ukrainian].
- Fraenkel, J., Wallen, N., & Hyun, H. (2012). How to design and evaluate research in education (8th ed.). McGraw-Hill.
- Honcharenko, S. (2010). Pedagogical research: Methodological advice for young scientists. Planer. [in Ukrainian].
- Horbachyk, A., & Salnikova, S. (2008). Analysis of sociological research data by means of SPSS: Study guide. Vezha, Lesya Ukrainka Volyn National University. [in Ukrainian].
- Klymeniuk, O. (2006). Technology of scientific research: Author's textbook. Aspekt-Polihrاف. [in Ukrainian].
- Mamchur, K. (2024). Methodology and methods of dissertation research on the problems of higher military educational institutions' faculty members' professional competence development. Bulletin of Science and Education, 5(23), 1109–1122. DOI: [https://doi.org/10.52058/2786-6165-2024-5\(23\)-1109-1122](https://doi.org/10.52058/2786-6165-2024-5(23)-1109-1122) [in Ukrainian].
- Mitina, S. (2024). Methods of mathematical statistics in psychological research: Study guidelines. TNU Publishing House. [in Ukrainian].
- NATO. (n.d.). Defence Education Enhancement Programme (DEEP). URL: https://www.nato.int/cps/en/natohq/topics_139182.htm
- NATO Standardization Office. (n.d.). NATO Standardization Agreement (STANAG) 6001, Edition 5: Language proficiency levels. Bureau for International Language Co-ordination. URL: <https://natobilc.org/stanag-6001/>
- Panasenko, E. (2015). Psychological-pedagogical experiment: Milestones of formation and development (second half of the 19th – early 20th century). Scientific Bulletin of V. O.

- Sukhomlynskyi Mykolaiv National University. Psychological Sciences, 2, 90–98. [in Ukrainian].
- Tverezovska, N., & Sydorenko, V. (2013). Methodology of pedagogical research: Study guide. Centre of Educational Literature. [in Ukrainian].
- Velychko, O., Kolomiets, L., Hordiienko, T., Shevtsov, A., Karpenko, S., & Haber, A. (2015). Group expert assessment and competence of experts (O. Velychko, Ed.). Bondarenko M. O. [in Ukrainian].
- Yagupov, V. (2019). Theory and methodology of military-pedagogical research: Textbook. Ivan Cherniakhovskyi National Defence University of Ukraine. [in Ukrainian].
- Zhosan, O. (Comp.). (2008). Pedagogical experiment: Study guide. Vasyl Sukhomlynskyi KOIPPO Publishing House. [in Ukrainian].



This is an open access journal and all published articles are licensed under a Creative Commons «Attribution» 4.0.