

УДК 378.14.014.13

Дар'я Проноза,
Національний університет оборони України
імені Івана Черняхівського, м. Київ
<https://orcid.org/0000-0002-7286-1366>
DOI: 10.33099/2617-1775/2022-01/197-206

АНАЛІЗ ТЕСТОВИХ ЗАВДАНЬ СТАНДАРТИЗОВАНИХ МОВНИХ ТЕСТІВ НА ВИЗНАЧЕННЯ РІВНЯ ВОЛОДІННЯ ІНОЗЕМНОЮ МОВОЮ ВІДПОВІДНО ДО ВИМОГ НАТО СТАНАГ 6001

В статті висвітлюється методика аналізу тестових завдань стандартизованих мовних тестів на визначення рівня володіння іноземною мовою, а саме визначення складності, індексу дискримінації тестового завдання та ефективності дистракторів. Аналіз дозволяє оцінити внесок кожного тестового завдання і якість тесту в цілому та вдосконалити їх.

Ключові слова: мовна підготовка; СТАНАГ 6001; мовне тестування; тестове завдання; дистракторний аналіз; складність; індекс дискримінації

Постановка проблеми. Володіння іноземними мовами особовим складом Збройних Сил (ЗС) України є важливою складовою їхнього реформування та необхідною умовою для міжнародного військового співробітництва й досягнення взаємосумісності зі збройними силами країн-членів НАТО. У зв'язку з цим актуальності набуває мовна підготовка та стандартизованість процедури мовного тестування відповідно до вимог НАТО СТАНАГ 6001. Рівень володіння іноземними мовами військовослужбовцями ЗС України оцінюється за допомогою стандартизованих мовних тестів, що розробляються згідно з вимогами стандартів НАТО СТАНАГ 6001 (мовний тест НАТО СТАНАГ 6001) для визначення рівня володіння іноземною мовою за чотирма видами мовленнєвої діяльності: аудіювання, говоріння, читання та письма. Щоб мовний тест надав якісні відомості щодо дійсного рівня володіння іноземною мовою тестуємих, тест та тестові завдання мають пройти випробування. Одним з видів випробувань є аналіз якості тестових завдань, за якого на основі результатів апробації тесту розраховуються його статистичні характеристики. Апробація тесту – це пробне тестування, що передує офіційному та проводиться з метою усунення помилок і недоліків у завданнях. Своєю чергою, визначення статистичних характеристик є головним засобом діагностики тестових властивостей завдання [7].

З огляду на зазначене вище, в контексті політики ЗС України щодо інтеграції до НАТО стає актуальним питання мовної підготовки та мовного тестування, що також посилюється Дорожньою картою, в якій зазначено вплив знання іноземної мови на кар'єрне зростання офіцерів.

Метою статті є висвітлення методики аналізу тестових завдань стандартизованих мовних тестів на визначення рівня володіння іноземною мовою відповідно до вимог НАТО СТАНАГ 6001.

Аналіз останніх досліджень і публікацій. Питаннями тестового контролю в освітньому процесі займалися багато вітчизняних та зарубіжних науковців.

Теоретико-методологічні аспекти розроблення мовних тестів та проведення мовного тестування висвітлені в працях В. С. Аванесова, І. Є. Булах, О. Г. Квасової, Л. О. Кухар, В. М. Кухаренко, А. Г'юза, В. Д. Крикун, Ю. В. Грищук, публікаціях міжнародних екзаменаційних центрів з іноземних мов, рекомендаціях експертів Бюро міжнародної мовної координації НАТО та інших. Однак, проблема методики оцінки якості тестових завдань як необхідного етапу в процесі розроблення та забезпечення валідності мовного тесту недостатньо повно висвітлюється в науковій літературі й залишається актуальною й сьогодні.

Виклад основного матеріалу. Стандартизовані мовні тести розробляються відповідно до визначених критеріїв, основними з яких є валідність та надійність. Українські дослідники розглядають валідність тесту як поняття, яке вказує на те, що тест вимірює і наскільки успішно він це вимірює. На думку науковців, поняття валідності використовується для загальної характеристики тесту щодо відповідності одержаних результатів меті та вимогам оцінювання.

Іншим критерієм, на який варто звернути увагу, є надійність методу вимірювання, що є мірою стійкості результатів, яка впливає на точність з якою можна виміряти ту чи іншу конкретну ознаку під час тестування. Перевірка надійності методу вимірювання стосується насамперед відновлення результатів при повторних вимірах. Тест вважається надійним, якщо він надає однакові показники для кожного тестуемого при повторному тестуванні [3].

Щоб тест був вдалим інструментом оцінювання, він має пройти апробацію, що є офіційним схваленням, винесеним на підставі випробування та перевірки з метою об'єктивної оцінки властивостей і якостей тестових завдань і тесту в цілому. Апробація є одним з важливих компонентів будь-якої системи тестування, зокрема тестування відповідно до вимог НАТО СТАНАГ 6001. Завдяки правильній організації апробації підвищуються показники якості тестування, а саме: валідність, надійність, об'єктивність та ефективність. Апробація тестових завдань проводиться, зокрема, для вирішення таких завдань: вхідний контроль коректності змісту тестових завдань; вхідний контроль наявності розрізнявальної здатності і вимірювальних властивостей тестових завдань; «калібрування» тестових завдань (класифікація завдань відповідно їх вимірювальних характеристик); перевірка та налагодження інструментарію оцінки [5; 9].

Аналіз якості тестових завдань на основі отриманих результатів апробації необхідний для вдосконалення як окремих тестових завдань, так і тесту в цілому. Аналіз дозволяє визначити складність тестового завдання (*facility value*), індекс дискримінації (*discrimination index*) тестового завдання та ефективність дистракторів (*distractor efficiency*) для завдань множинного вибору [8]. Аналіз тестових завдань емпіричними (статистичними, математичними) методами дозволяє одержати інформацію про їх приховані дефекти, що не можуть бути виявлені за допомогою аналітичних (експертних) методів.

Згідно з рекомендаціями експертів Бюро міжнародної мовної координації

НАТО [8], аналіз якості тестових завдань починається з підрахунку балів за виконання усіх тестових завдань тесту. Далі слід розташувати бали в порядку від найвищого до найнижчого та розділити їх на три рівні групи: найвищі бали, найнижчі бали та середні бали. Аналіз тестових завдань зазвичай включає тільки групи з найвищими та найнижчими балами. Для кожного тестового завдання необхідно створити таблицю, подібну до наведеної нижче, на якій позначається значення складності, значення індексу дискримінації, ефективність дистракторів тестового завдання та вказується ключ (правильна відповідь на завдання).

Значення складності тестового завдання (*facility value, FV*) – це відсоток тестуємих, які правильно відповіли на завдання. Чим більше відсоток, тим легше завдання. Пропорція для тестового завдання зазвичай позначається як p . Це відносна частота, з якою тестуємі вибрали правильну відповідь. Завдання, на яке правильно відповіли 85% тестуємих, має значення $p = 0,85$, а завдання, на яке правильно відповіли 20% тестуємих, відповідно має значення $p = 0,20$. Чим вище значення p , тим простіше завдання для даної вибірки тестуємих. Наприклад, якщо лише 20% тестуємих виконали завдання, його можна вважати складним для даної вибірки, якщо 80% – легким. Залежно від значення складності тестове завдання кваліфікується таким чином:

- $0,20 < p < 0,36$ – надто складне;
- $0,36 < p < 0,84$ – середньої складності;
- $0,84 < p$ – надто легке.

За цим показником просто диференціювати легке, середньої складності, складне або відгадане завдання [2]. Добір завдань за показниками складності важливий для успішного використання тестів. Так, при доборі надто складних тестових завдань валідність і надійність тесту різко зменшуються. Надто прості тестові завдання призведуть до одноманітності тесту та його неефективності [4]. Приклад аналізу складності завдання наведено в таблиці 1.

ЗАВДАННЯ № 10 - 35 тестуємих вибрали правильну відповідь (С).

Таблиця 1

| | Варіанти відповідей | | | | | Індекси | |
|---------------|---------------------|---|----|---|-----------|------------|---------------|
| | A | B | C | D | Пропущено | Складність | Дискримінація |
| Верхня група | 2 | 0 | 17 | 1 | 0 | 58% | 0,55 |
| Середня група | 4 | 1 | 12 | 3 | 0 | | |
| Нижня група | 5 | 4 | 6 | 3 | 2 | | |

35 з 60 тестуємих відповіли правильно. Це становить 58%. Це і є значення складності або p .

Чим вище відсоток тестуємих, що правильно відповіли на завдання, тим воно легше. Значення складності завдання характеризує як саме тестове завдання, так і рівень підготовки тестуємих. Тестове завдання, яке є легким для

тестуємих з високим рівнем знань, ймовірно, не буде легким для тестуємих з низьким рівнем знань. Значення складності сильно впливає як на варіативність результатів тестування, так і на точність, з якою тестові завдання диференціюють між тестуємими. Якщо завдання дуже легке, ви можете отримати p рівне 1,0 (всі відповіді правильно); якщо завдання занадто складне, ви можете отримати p рівне 0,00 (ніхто не відповів правильно). В обох випадках варіативність відсутня, що означає, що завдання не диференціює між тестуємими.

Індекс дискримінації (*discrimination value/discrimination index, DI*) або розподільна здатність тестового завдання – це коефіцієнт кореляції між балом за весь тест і балом за окреме тестове завдання. Якщо тест і окреме тестове завдання вимірюють одне й те саме, слід очікувати, що тестуємі, які добре склали тест, дадуть правильну відповідь на завдання, а ті, хто погано склали тест, дадуть неправильну відповідь на завдання. Хороше тестове завдання розрізняє між тестуємими, які набрали високі бали на тесті, і між тими, хто набрав низькі бали. Чим вище індекс дискримінації, тим краще тестове завдання, оскільки це значення вказує на те, що завдання диференціює верхню групу, яка має дати більше правильних відповідей на завдання, від нижньої групи. Якщо всі відповідають на питання правильно чи неправильно, завдання не диференціює між тестуємими, тому його слід вилучити або переробити. В цьому випадку, значення дискримінації становить 0,00. Якщо усі тестуємі в нижній групі і ніхто у верхній відповів правильно на завдання, то таке завдання поводить протилежно решті завдань тесту. Швидше за все, таке завдання є неоднозначним або має недоліки.

Критерії індексу дискримінації тестових завдань наведені в таблиці 2 [6].

Критерії індексу дискримінації тестових завдань

Таблиця 2

| Значення індексу дискримінації, % | Висновок |
|-----------------------------------|---|
| Понад 50 | Нормальна дискримінація |
| 30-50 | Можливе незначне коригування завдання |
| 20-29 | Завдання потрібно переглянути та переробити |
| 0-19 | Завдання потрібно вилучити або суттєво переробити |
| Від'ємне значення | Завдання потрібно вилучити з тесту |

Дискримінація є важливим показником, оскільки чим краще завдання розрізняє між тестуємими, тим надійнішим буде тест.

Індекс дискримінації (*DI*) може бути розраховано різними способами. Ми пропонуємо використовувати такі формули:

1. Відніміть значення складності завдань для нижньої групи від значення складності завдань для верхньої групи:

$$DI = FV_{upper} - FV_{lower}$$

2. Відніміть кількість тестуємих в нижній групі, що вибрали правильну відповідь, від кількості тестуємих у верхній групі, що вибрали правильну відповідь, і розділіть на кількість тестуємих у верхній групі:

$$DI = \frac{N_{upper} - N_{lower}}{total N_{upper}}$$

Слід звернути увагу на те, що значення індексу дискримінації не свідчить про валідність завдання. Індекс дискримінації свідчить про внутрішню послідовність, тобто про те, наскільки кожне завдання вимірює те, що вимірює й весь тест. Такі дані аналізу тестових завдань можна інтерпретувати як доказ валідності тестових завдань лише в тому випадку, якщо валідність всього тесту була доведена або може бути обґрунтовано припущена. Також слід зазначити, що низький індекс дискримінації не обов'язково свідчить про дефекти завдання. Слід уважно проаналізувати завдання, щоб зрозуміти, чи воно не вимірює щось інше, аніж решта тесту.

Тестові завдання множинного вибору складаються з ключа (правильної відповіді) та дистракторів (неправильних варіантів відповідей). Дистрактор (від англ. to distract - відволікати) – це неправильна, але правдоподібна відповідь, що використовується в процесі тестування для відволікання уваги від правильної відповіді тих тестуємих, які або зовсім не знають правильну відповідь, або, маючи дуже туманні уявлення про предмет тестування, намагаються вгадати її під час тестування. Від дистракторів залежить якість окремого завдання та тесту в цілому. Чим більше неякісних дистракторів у завданнях, тим гірший тест. Зробити неправильні відповіді правдоподібними – одне з важливих завдань розробника тесту.

Дистракторним аналізом називається теоретико-емпіричне дослідження правильності та ефективності підбору всіх відповідей, що додаються до кожного завдання тесту. Теоретична частина аналізу стосується питань змісту та форми завдань. Емпірична частина стосується організації апробації з метою визначення властивостей завдань та властивостей відповідей. Дистракторний аналіз здійснюється для визначення відносної корисності або ефективності дистракторів у кожному завданні.

Якщо дистрактор не вибирає жоден тестуємий, це може означати, що він не має властивостей дистрактора. Тому його необхідно або переформулювати та випробувати наново, або замінити іншим дистрактором.

Дистрактор можна вважати ефективним, якщо його вибирає значна частина тестуємих, які мають низьку підготовленість. Це може означати, що він добре диференціює тестуємих за рівнем підготовленості. Якщо дистрактор виявляється привабливим для тестуємих з високим рівнем підготовленості, це небажана властивість дистрактора. Такий дистрактор може бути неоднозначним, тому його слід переробити, щоб він був більш однозначно неправильним.

Найнижчий поріг вибору для дистрактора зазвичай встановлюється в 5 відсотків. Якщо дистрактор нікого не привабив, він може бути занадто очевидно неправильним. Тому він є слабким і не виконує свою роль в процесі

тестування.

Якщо якийсь дистрактор притягує до себе дуже великий відсоток тестуємих, 80% і більше, то такий дистрактор також краще замінити, оскільки інші дистрактори при цьому фактично не працюють. [1]

Аналіз ефективності дистракторів проводиться шляхом визначення відсотка тестуємих, які вибрали кожен дистрактор:

$$D \text{ efficiency} = \frac{N_d}{N_t}$$

N_d = кількість тестуємих, які вибрали конкретний дистрактор

N_t = кількість тестуємих, які вирішували завдання

Слід звернути увагу на те, що вибірка учасників для апробації повинна бути якомога більш репрезентативною, оскільки дистракторний аналіз завдань на основі невеликих вибірок є досить умовним.

Приклади аналізу завдань наведені у таблиці 3, таблиці 4 та таблиці 5.

Завдання № 1 – правильна відповідь А

Таблиця 3

| | Варіанти відповідей | | | | | Індекси | |
|---------------|---------------------|---|---|---|-----------|------------|---------------|
| | А | В | С | Д | Пропущено | Складність | Дискримінація |
| Верхня група | 8 | 2 | 0 | 0 | 0 | 70% | 0,20 |
| Середня група | 7 | 3 | 0 | 0 | 0 | | |
| Нижня група | 6 | 3 | 0 | 0 | 1 | | |

Давайте проаналізуємо вище зазначений приклад.

- Завдання досить просте, оскільки 70% тестуємих відповіли на нього правильно (0,70).

- Завдання погано розрізняє між верхньою та нижньою групами (значення індексу дискримінації = 0,20)

- З огляду на ефективність дистракторів, ми бачимо, що варіанти відповідей «С» і «Д» нікого не привабили, що свідчить про те, що дистрактори слабкі (можливо, занадто очевидно неправильні).

ЗАВДАННЯ № 2 – правильна відповідь В

Таблиця 4

| | Варіанти відповідей | | | | | Індекси | |
|---------------|---------------------|---|---|---|-----------|------------|---------------|
| | А | В | С | Д | Пропущено | Складність | Дискримінація |
| Верхня група | 0 | 7 | 1 | 2 | 0 | 47% | 0,50 |
| Середня група | 1 | 5 | 2 | 2 | 0 | | |
| Нижня група | 2 | 2 | 2 | 2 | 2 | | |

Давайте проаналізуємо приклад з таблиці 4.

- Це завдання відповідає меті тесту. Значення складності завдання становить 0,47, що означає, що 47% учасників усієї групи відповіли правильно.
- Завдання добре розрізняє між тестуємими і має хорошу дискримінаційну здатність (0,50), що означає, що 50% тестуємих відповіли правильно, а 50% - неправильно. Це ідеальне завдання.
- З точки зору ефективності дистракторів, це також ідеальне завдання, оскільки дистрактори привабили однакову кількість тестуємих.

ЗАВДАННЯ №3 – правильна відповідь С

Таблиця 5

| | Варіанти відповідей | | | | | Індекси | |
|---------------|---------------------|---|---|---|-----------|------------|---------------|
| | А | В | С | Д | Пропущено | Складність | Дискримінація |
| Верхня група | 3 | 3 | 2 | 2 | 0 | 33% | 0,30 |
| Середня група | 2 | 3 | 3 | 1 | 1 | | |
| Нижня група | 1 | 2 | 5 | 1 | 1 | | |

Давайте проаналізуємо цей приклад.

- Оскільки лише 33% учасників усієї групи правильно відповіли на це завдання, його значення складності не є високим.
- Однак ми бачимо, що завдання не поводить ся так, як решта завдань тесту, тому що нижня група вибрала правильну відповідь, але не верхня.
- Усі дистрактори привабили однакову кількість тестуємих. Однак ключ (правильна відповідь) не привабив верхню групу, що свідчить про те, що правильна відповідь може бути неоднозначною [8].

Кожен вид завдання має власні особливості в аналізі. Проте, в будь-якому випадку можна проаналізувати відповіді, які дають тестуємі. Варіанти відповідей, які не вибирає ніхто, слід вилучати. В завданнях множинного вибору всі варіанти відповідей повинні бути схожими на правильну відповідь і не бути відверто неправильними. Якщо ж усі тестуємі вибирають лише правильну відповідь, то таке завдання є занадто легким і його необхідно вилучити або замінити всі дистрактори.

Висновки. У статті обґрунтовано і розкрито зміст показників якості тестових завдань, таких як складність тестового завдання, індекс дискримінації тестового завдання та ефективність дистракторів, які застосовуються для аналізу тестових завдань. Наведена методика дозволяє опрацьовувати результати апробації з можливістю аналізу, оцінки якості та удосконалення кожного тестового завдання, що є необхідною умовою для забезпечення валідності та надійності тесту.

Напрями подальших досліджень. Напрями подальших досліджень аналізу тестових завдань для стандартизованих мовних тестів на визначення рівня володіння іноземною мовою відповідно до вимог НАТО СТАНАГ 6001

ми вбачаємо у аналізі вбудованих систем аналізу завдань у програмах комп'ютерного мовного тестування та методиці статистичного аналізу та обробки результатів тестування.

ЛІТЕРАТУРА

1. Аванесов В. Дистракторный анализ. – Режим доступу: <http://testolog.narod.ru/Theory46.html>
2. Булах І. Є. Створюємо якісний тест: навч. посіб. / І. Є. Булах, М. Р. Мруга. – К. : Майстер-клас, 2006. – 160 с.
3. Гришук Ю. В., Галкіна В. Д. Забезпечення валідності та надійності мовних тестів згідно вимог стандартів НАТО «STANAG 6001». – Режим доступу: https://ela.kpi.ua/bitstream/123456789/17374/1/17_Hryshchuk.pdf
4. Конструювання тестів. Курс лекцій: навч. посіб. / Л.О. Кухар, В.П. Сергієнко. – Луцьк, 2010. – 182 с.
5. Методика проведення апробації стандартизованого мовного тесту на визначення рівня володіння іноземною мовою: навчальний посібник / колектив авторів; за заг. ред. к. військ. н. С. С. Невгада. К.: НУОУ ім. Івана Черняхівського, 2021. 38 с.
6. Мокрієв М. В. Аналіз тестових завдань засобами Moodle. – Режим доступу: <http://2017.moodlemoot.in.ua/course/view.php?id=83&lang=en>
7. Сергієнко В. П., Кухар Л. О., Галицький О. В., Микитенко П. В. Використання вбудованої системи аналізу тестових завдань в LCMS MOODLE. // Інформаційні технології і засоби навчання, 2014, Том 41, №3. – Режим доступу: https://ino.nau.edu.ua/files/repository/articles/ITZN_2014_41_3_21.pdf
8. BILC Language Testing Seminar, 3-14 June, 2019.
9. Gawliczek, P. ., Krykun, V., Tarasenko, N., Tyshchenko, M., & Shapran, O. (2021). Computer Adaptive Language Testing According to NATO STANAG 6001 Requirements. Advanced Education (2021): 19-26.

REFERENCES

1. Avanesov V. Distractor analysis. - Access mode: <http://testolog.narod.ru/Theory46.html>
2. Bulakh I. E. Creating a quality test: textbook./ I. E. Bulakh, M. R. Mruga. – K.: Master-class, 2006. - 160 p.
3. Grishchuk Y. V., Galkina V. D. Ensuring the validity and reliability of language tests in accordance with NATO standards STANAG 6001. - Access mode: https://ela.kpi.ua/bitstream/123456789/17374/1/17_Hryshchuk.pdf
4. Designing tests. Course of lectures: textbook. / L. O. Kukhar, V. P. Sergienko. - Lutsk, 2010. - 182 p.
5. Methods of piloting a standardized language proficiency test: textbook / team of authors; edited by S. S. Nevhad PhD. K.: NDU named after Ivan Chernyakhovsky, 2021. 38 p.
6. Mokriev M. V. Analysis of test tasks using Moodle. - Access mode: <http://2017.moodlemoot.in.ua/course/view.php?id=83&lang=en>
7. Sergienko V. P., Kukhar L. O., Galitsky O. V., Mykytenko P. V. Using the built-in system of analysis of test tasks in MOODLE. // Information technologies and teaching aids, 2014, Volume 41, №3. - Access mode: https://ino.nau.edu.ua/files/repository/articles/ITZN_2014_41_3_21.pdf
8. BILC Language Testing Seminar, 3-14 June, 2019.
9. Gawliczek, P. ., Krykun, V., Tarasenko, N., Tyshchenko, M., & Shapran, O. (2021). Computer Adaptive Language Testing According to NATO STANAG 6001 Requirements. Advanced Education (2021): 19-26.

РЕЗЮМЕ

Дарья Проноза,
Национальный университет обороны Украины
имени Ивана Черняховского, г. Киев

Анализ тестовых заданий стандартизированных языковых тестов на определение уровня владения иностранным языком в соответствии с требованиями НАТО СТАНАГ 6001

В статье освещается методика анализа тестовых заданий стандартизированных языковых тестов на определение уровня владения иностранным языком, а именно определение сложности, индекса дискриминации тестового задания и эффективности дистракторов. Анализ позволяет оценить вклад каждого тестового задания и качество теста в целом и усовершенствовать их.

Ключевые слова: языковая подготовка; СТАНАГ 6001; языковое тестирование; тестовое задание; дистракторный анализ; сложность; индекс дискриминации

SUMMARY

Daria Pronoza,
National Defence University of Ukraine
named after Ivan Cherniakhovskiy, Kyiv

Analysis of test items of standardized language proficiency tests in accordance with NATO STANAG 6001

The article examines a methodology of analyzing test items of standardized language proficiency tests in accordance with the NATO STANAG 6001, namely the facility value, discrimination index and distractor efficiency of test items. Item analysis is conducted to evaluate the contribution of each item and the quality of the overall test. Results of item analysis are used to improve test items, which in turn ensures test validity and reliability.

In order for a language test to provide accurate information about the level of language proficiency of test takers, test items must be pre-tested and piloted. Piloting is a trial test the results of which are used to eliminate errors and inaccuracies in the test items. Calculation of statistical features is the primary means of diagnosing test item parameters.

The facility value of a test item is the percentage of test takers who answered the item correctly. The larger the percentage, the easier the item. The selection of items according to facility value is important for the successful use of tests. Selecting test items that are too difficult drastically reduces test validity and reliability. Items that are too easy will lead to test monotony and ineffectiveness.

If the test and a separate item measure the same thing, test takers who do well on the test are expected to answer that item correctly, and those who do not do well on the test are expected to answer the item incorrectly. A good item must discriminate between test takers who score high on the whole test and those who score low. If everybody answers the item correctly or incorrectly, the item does not discriminate and should be removed or rewritten.

Distractor is an incorrect alternative in a multiple choice item used in the testing process to divert attention from the correct answer of those test takers who either do not know the

correct answer at all or have very vague ideas about the subject of testing and try to guess the correct answer during testing. The quality of a separate item and of the overall test depends on distractors. Distractor analysis is important in determining the relative usefulness or efficiency of the decoys in each item. If nobody selects a particular distractor, it should be rewritten or changed altogether. Distractor analysis is done by looking at the percentage of test takers who choose each distractor.

The prospect of further research in the theoretical framework is determined by the authors in analyzing built-in systems of test item analysis in computer language testing programs and methods of estimating, describing, and analyzing the test results.

Key words: *language training; STANAG 6001; language testing; test item; distractor analysis; facility value; discrimination index*