УДК 37.091.279.2:005.963.1

**Бурак Мар'яна Ігорівна**
аспірант кафедри англійської філології
Тернопільський національний педагогічний університет імені В. Гнатюка
м.Тернопіль, Україна

# СПОСОБИ МІНІМІЗАЦІЇ ВПЛИВУ РЕЙТЕРІВ НА ВАЛІДНІСТЬ ТА НАДІЙНІСТЬ ОЦІНЮВАННЯ МОВНИХ МАСШТАБНИХ СТАНДАРТИЗОВАНИХ ТЕСТІВ

*Стаття розглядає основні фактори впливу рейтерів на валідність та надійність результатів суб'єктивного оцінювання продуктивних видів іншомовної комунікативної компетентності, розглядає різні аспекти тестування як найпоширенішого засобу визначення рівня володіння мовою та аналізує дослідження ефективності навчальних програм для рейтерів з метою мінімізації неточності рейтерів та їхньої непослідовності в оцінюванні, особливо важливої у випадку масштабних стандартизованих тестів.*

*Ключові слова: тестування, валідність, надійність, послідовність, рейтер, іншомовна компетентність, оцінювання*

**Вступ**. Дослідження шляхів оптимізації методів та засобів мовного оцінювання останнім часом привертають увагу науковців не лише за кордоном, але й у вітчизняній прикладній лінгвістиці та педагогіці. Це зумовлено, зокрема, інтенсивними процесами глобалізації, академічної та робочої міграції українців, розширенням міжнародних суспільно-економічних зв'язків нашої країни та науково-дослідної співпраці українських вчених з колегами чи навчальними закладами за кордоном. Такі тенденції у суспільному житті вимагають від українців певного рівня володіння іноземними мовами та часто його підтвердження сертифікатом про складання відповідного іспиту. Так існує низка популярних нині засобів визначення рівня сформованості іншомовної компетентності з англійської мови, на зразок міжнародних іспитів TOEFL, TOEIC, IELTS, Cambridge English Test, CaMLA, OPI тощо, проте, об'єктивність їхніх результатів усе ще викликає питання.

**Аналіз останніх досліджень і публікацій.** Уже більше як століття науковці світу ведуть дискусії про найефективніші засоби перевірки досягнень іншомовної компетентності (Л.Бахман, Д.Барвелл, Р.Вуд, Ч.Гандшін, М.Канейл, Ф.Келлі, Ч.Олдерсон, А.Палмер, Ч.Старч, М.Свейн, Ґ.Фулхер тощо). Проте одним із оптимальних засобів такого оцінювання на сьогодні все ще залишається мовне тестування. Розробка та використання тетсів з метою оцінювання мовної компетентності − сфера досліджень лігводидактики, яка виникла та вже широко розвинулася за кордоном. Перші наукові дослідження з педагогічної тестології (науки про тести) з'явилися на початку XX століття у США і стосувалися розробки та використання тестів як засобу об'єктивного контролю рівня знань. Відтак за минулі сто років американська та західна тестологія кілька разів кардинально змінювала акценти, перебувала у постійному розвитку та пошуку оптимальних методів оцінювання, розробила систему понять, виробила методологію і стала однією з провідних сфер прикладної лінгвістики та галузі педагогічних вимірювань. Тимчасом в Україні, як зазначав у 2009 р. директор Центру оцінювання освіти І.Лікарчук в одному із своїх інтерв'ю, "немає власної школи тестування. Взагалі, в Україні немає такої науки як тестологія, немає такої науки як педагогічні вимірювання. Те, що дуже давно є в багатьох країнах світу, у нас тільки починає розвиватися" [1, с.88]. Однак потрібно взяти до уваги, що наукові дослідження вітчизняних вчених останніх років свідчать про певні зрушеннння у цій сфері (І.Булах, Л.Морська, С.Ніколаєва, Л.Петращук, О.Петренко тощо), про зародження української тестології і її поступовий розвиток.

Огляд наукових досліджень українських авторів, втім, показав, що увага вітчизняних науковців здебільшого зосереджена на дидактичних принципах контролю знань, які реалізуються у змісті, формі та методах його проведення. Втім один із аспектів перевірки, який має суттєвий, а іноді і визначальний, вплив на результат контролю − професійні знання та вміння екзаменаторів (рейтерів), які власне оцінюють відповіді тестованих, − залишається поза їхньою увагою, тоді як у закордонній тестології питання професійної підготовки фахівців з оцінювання рівня сформованості іншомовної компетентності набуває все більшого інтересу (С.Вейґл, А.Кемпбелл, Т.Макнамара, Д.Чарні тощо).

На нашу думку, сьогодні для України питання підготовки оцінювачів (рейтерів) рівня сформованості комунікативної компетентності є особливо актуальним. Сучасне масове використання стандартизованих масштабних тестів з високим статусом у вигляді Зовнішнього незалежного оцінювання, проведення атестації осіб, які претендують на вступ на державну службу, щодо вільного володіння державною мовою (Постанова Кабінету міністрів України № 301 від 26.04.2017) та гостра потреба у розробці та проведенні стандартизованих тестів на визначення рівня володіння українською мовою як іноземною вимагають негайного вивчення проблеми для ретельної професійної підготовки екзаменаторів/рейтерів, особливо при оцінюванні продуктивних (творчих) типів завдань з говоріння та/чи письма.

**Метою** статті є аналіз впливу рейтерів на оцінювання мовного масштабного стандартизованого тесту та пошук способів його мінімізації. Завдання - визначити ефективність курсів підготовки/перепідготовки рейтерів у підвищенні рівня валідності та надійності оцінювання, а відтак зменшенні впливу рейтера на результати тестування.

**Виклад основного матеріалу.** Найоптимальнішим засобом перевірки сформованості іншомовної компетентності вважають мовне або лінгводидактичне тестування. Проте науковці не досягли одностайності у визначенні поняття "тесту"/ "тестування". Так, В.Коккота вважає, що тест - це "короткочасне, технічно просто обставлене випробування, яке проводиться в рівних для всіх умовах і яке має вигляд такого завдання, вирішення якого піддається кількісному обрахунку" [2, с.7] Власне лінгводидактичним тестом, який і є об'єктом нашого дослідження, Н.Гарматюк та В.Марценюк [3, с.19] влучно вважають "підготовлений до певних вимог комплекс завдань, які попередньо випробовувались з метою встановлення їх якості і які дозволяють виявити в учасників тестування їх рівень лінгвістичної або комунікативної компетенції та оцінити результати тес-

© Бурак М.І.

тування за заздалегідь виведеними критеріями".

Оскільки предметом нашого дослідження є професійна діяльність рейтера в оцінюванні результатів тесту, то, вважаємо, що "тестування" у цьому контексті слід розглядати не лише як процедуру проведення контролю у формі тесту, а у широкому значенні цього терміна у трактуванні С.Ніколаєвої "як сукупності процедурних етапів планування, складання й випробування тестів, обробки та інтерпретації результатів проведення тесту" [2, с.4] включно з професійною підготовкою укладачів та оцінювачів самих тестів.

Стандартизовані тести з високим статусом (селективні тести з життєво важливими наслідками для тестованих, на зразок можливості навчання у певному закладі, визначення професійної придатності чи відповідності посаді, шансів на отримання громадянства тощо), які є об'єктом нашого дослідження, – це ті процедури оцінювання, результати яких можуть мати доленосне значення для кандидатів, а відтак мають бути якомога більше націлені на валідність як самих тестів, так і їхнього оцінювання.

Масштабне стандартизоване тестування – важке і комплексне. Його можна вважати видом комплексного контролю, проте, не в розумінні Ю.Головач [4, с.3], яка визначає такий вид перевірки як сукупність "пооопераційного поточного та періодичного контролю" на різних етапах формування мовленнєвих навичок, а як формування цілісного уявлення про рівень сформованості цих навичок на основі "тестів, які пройшли попереднє випробування на великій кількості учасників і мають кількісні показники якості" [3, с.20].

Сучасні масштабні мовні тести з високим статусом є здебільшого багатокомпонентними та багаторівневими. Вони передбачають перевірку усіх видів мовленнєвої діяльності на репродуктивному, конструктивному та творчому рівнях. Часто такі тести поєднують репродуктивні (читання та аудіювання) та продуктивні (говоріння та письмо) види у межах одного завдання. Такі інтегровані завдання дозволяють занурити кандидата у мовленнєву ситуацію, максимально наближену до реального життя, де комуніканти сукупно оперують різними видами мовлення. Саме такий тип завдань, на думку Кладкової [4, с.81], є оптимальним для перевірки загального рівня володіння мовою та "якнайкраще корелює з основними принципами компетентісного підходу" до викладання іноземних мов, який передбачає, на думку С.Ніколаєвої, "формування у студентів здатності застосовувати набуті знання, навички і вміння у конкретних ситуаціях" [4, с.81]. Проте процес створення та валідації таких тестів важкий та дорогий, а завдання, хоч і наближені до реальних життєвих умов, все ж штучно створені або й опосередковані, у випадку парного оцінювання усного мовлення за участю інтерв'юера (інтерлокутора). І саме такі продуктивні, творчі або інтегровані типи завдань передбачають суб'єктивне оцінювання їхніх результатів, яке базується на судженні того, хто перевіряє, а отже вимагає від нього особливих знань та вмінь.

Цілком виправдано, що більшість рейтерів провідних тестових компаній з англійської мови як іноземної перед початком своєї професійної діяльності зобов'язані пройти короткий курс навчання під керівництвом провідного фахівця або кваліфікованого працівника тестового центру для чіткого розуміння критеріїв, типів зразків та самої процедури оцінювання. Згодом рейтер повинен представити на розгляд акредитаційної комісії певну кількість своїх власних пробних оцінювань. Лише після підтвердження того, що відповідний відсоток їхніх оцінювань був надійним та валідним рейтер може отримати сертифікат про акредитацію та офіційно працювати. Проте часто така акредитація є тимчасовою та вимагає від рейтера повторного підтвердження професійності: Впродовж усього терміну своєї роботи оцінювач зобов'язаний проходити систематичні вишколи з так званої рестандартизації або поточної стандартизації [5, с.91].

Рейтери, як звичайні люди, відрізняються мірою своєї строгості чи поблажливості в оцінюванні. Адже не існує бездоганних рейтерів і традиційно в дослідженнях розрізняють поняття ідеалізованого (ідеального) рейтера та фактичного (звичайного) рейтера. Дослідник категорії оцінки з точки зору логіки О.Івін [6, с.9] вважає, що "не лише кожна людина по-різному оцінює одне й теж явище, але й ставлення однієї особи до певного об'єкта може змінюватися в діахронічному розвитку" (наприклад, при повторному оцінюванні того самого зразка через деякий час, рейтер може змінити свою думку щодо результату). Відтак, очевидно, що такий притаманний рейтерам певний особистий і ситуативний рівень строгості/поблажливості чи упередженості до тих чи інших завдань або кандидатів має безпосередній вплив на об'єктивність результатів. До того ж оцінювання продуктивних видів мовленнєвої діяльності як з рідної, так і з іноземної мови є цілісним, а тому не завжди точним. Аби мінімізувати вплив таких факторів на валідність результатів тестування і виникає необхідність у ретельній підготовці рейтерів та постійному відслідковуванні їхньої роботи. З іншого боку, строгість чи поблажливість рейтера є його постійною психологічною рисою, яку неможливо калібрувати відповідно до вимог надійного оцінювання за допомогою будь-яких навчальних методів чи засобів. І хоч існують нечисельні докази того, що професійна підготовка може дещо зменшити суттєву різницю в строгості оцінювання рейтерів (С.Вейгл) [7, с.2], вони потребують ще додаткового вивчення та підтвердження. Водночас, оскільки дієвим способом мінімізувати вплив різного ступеня строгості оцінювання є додатковий аналіз результатів тестування за допомогою метричної моделі Джорджа Раша, то зменшення відмінностей у строгості оцінювання може і не бути основною метою навчання рейтерів. Адже, як вважає ряд дослідників, ефект строгості рейтера, який полягає у послідовному оцінюванні кандидатів одним рейтером або надто строго, або надто поблажливо у порівнянні з іншими, − неєдиний систематичний вплив оцінювачів на визначення рівня знань та вмінь (Т.Макнамара) [5, с.91]. Серед інших таких факторів впливу (Т.Макнамара) [8, с.4], зокрема, називає: *непослідовність* − схильність рейтера до непослідовного застосування тієї чи іншої категорії шкали оцінювання у порівнянні з іншими рейтерами; *ефект ореолу* − неспроможність рейтера відділяти окремі концептуальні категорії при оцінюванні, а відтак присуджувати подібні результати за різними категоріями шкали оцінювання, керуючись загальним враженням про кандидата; *ефект доцентровості* − тяжіння до виставлення середніх оцінок та уникання надто високих чи надто низьких результатів; *неточність (упередженість)* − схильність до надто строгого чи поблажливого оцінювання одного з аспектів тестової ситуації (скажімо, певної категорії шкали оцінювання чи певного завдання).

Одним із способів уникнути ненадійність в оцінюванні є його максимальна відповідність стандартам, описаним у рубриках оцінювання. «Стандартизуючи процес оцінювання, рубрики підвищують його надійність. Кожне завдання оцінюють за тими самими критеріями, а виконання інструкцій до рубрик мінімізує

вплив неточностей» (О.Гітт і Е.Гельмс) [9, с.117] Інші дослідники теж говорять про важливість рубрик, які «зменшують неточність оцінювання" (А.Пауелл) [5, с.117], «підвищують рівень послідовності та надійності оцінювання» (А.Кемпбелл) [9, с.117] та «зосереджують на конкретних критеріях функціонування» (С.Райтмайер) [9, с.117]. Відтак, очевидним є те, що попереднє ознайомлення нових рейтерів з рубриками та шкалою оцінювання, а також їх систематичний перегляд уже сертифікованими екзаменаторами, повинен бути і, зазвичай, є частиною навчальних програм для рейтерів. Доцільність таких повторень підтверджують і відгуки самих рейтерів на одну із навчальних програм [8, с.16]. Однак, як свідчать нещодавні дослідження М.Ловорн та А.Резаї [9, с.118], лише використання рубрик під час вишколів рейтерів автоматично не підвищить міжрейтерську надійність чи точність результату. Для цього потрібно також тренувати екзаменаторів практично оцінювати відповіді (Дж.Макклеллан) [9, с.118].

Запорукою ефективного тестування є відповідність тестів якісним характеристикам, основними серед яких Грищук та Галкіна [10, с.112] вважають валідність (відповідність тесту меті оцінювання та його ефективність в її досягненні) та надійність (рівень співпадіння результатів двох чи більше тестів для оцінювання тих самих характеристик тестованих (К.Ікеда) [11, с.4]. Обидві ці характеристики є "суто вимірювальними" (Овчарук) [12, с.67], тобто можуть статистично обчислюватися, а відтак є максимально репрезентативними. Впродовж останніх років із стрімким розвитком тестології все більше науковців говорять про важливість підготовки рейтерів власне задля підвищення надійності їхнього оцінювання (П.Дідріх, С.Карлтон, Дж.Френч) [7, с.2], що передбачає послідовність в оцінюванні того самого продукту чи функції двома чи більше незалежними рейтерами. Точність такого оцінювання зазвичай обчислюється шляхом порівняння особистих результатів окремих рейтерів з «правдивим» результатом експерта. С.Стемлер [9, с.118] запропонував термін «консенсусна оцінка» для позначення ступеня схожості оцінок того самого матеріалу різними екзаменаторами. Така величина обчислюється як відсоток загальної узгодженості та/чи відсоток суміжної узгодженості (в межах одного рівня) (А.Джонсон і Дж.Свінгбі) [9, с.118]. Надійність оцінювання забезпечується різними рівнями процесу стандартизації, який передує самій процедурі тестування - це і визначення шкали та процедури оцінювання, і намагання досягти подібності в оцінюванні та узгодженості у результатах оцінювання рейтерів.

Однак, як стверджував Дж.Лінакр [7, с.2], «функція підготовки рейтерів полягає не в примусі їх дійти згоди між собою (міжрейтерської надійності), а радше у навчанні їх власній послідовності (внутрішньорейтерській надійності)». Адже доведено (П.Сток і Дж.Робінсон) [7, с.2], що незначні розбіжності в оцінюваннях між рейтерами, які існують навіть після навчання, є природними та невідворотними. Тому основною метою підготовки є «застерегти рейтерів від винесення власних вердиктів» (Д.Чарні) та «підвищити загальну послідовність шляхом підвищення внутрішньорейтерської послідовності» (Т.Макнамара) [7, с.2] за рахунок зменшення індивідуальних неточностей рейтерів. З іншого боку, кількісне напівекспериментальне дослідження іранських науковців М.Фахім та Г.Біджані довело ефективність навчальних програм для рейтерів у плані підвищення рівня узгодженості їхнього оцінювання, а відтак у збільшенні міжрейтерської надійності. І. Шепес також вважає очевидним, що вишколи рейтерів є обов'язковою передумовою досягнення максимальної міжрейтерської та внутрішньорейтерської послідовності [13, с.47].

Про доцільність навчання рейтерів говорили і С.Майер та Дж.Штульман [9, с.134]. Вони пропонували включити до курсу підготовки: 1. Зразки відповідей кожного рівня з можливістю обговорення з експертами обґрунтування оцінок. Вивчення зразків потрібно починати з чітких та однозначних прикладів конкретного рівня виконання і лише згодом варто додавати на розгляд інші нетипові види відповідей (Дж. Макклеллан) [9, с.118]. 2. Можливість вправляння в оцінюванні та в подальшому зіставлення результатів з оцінками експертів. 3. «Калібрування» рейтерів, яке полягає у продовженні практичного оцінювання з рецензією, допоки оцінювання рейтера не буде послідовно узгоджуватися з попередньою оцінкою експертів.

Першочерговим ж елементом навчання, як уже зазначалося, повинно стати все ж первинне/повторне вивчення критеріїв оцінювання [14, с.38; 8, с.16].

Отож, як показали численні дослідження С.Вейгл [8, с.5], П.Макінтайер і Ґ.Вігглсворт [14, с.38], вишкіл рейтерів є ефективним, оскільки усуває суттєву відмінність у строгості оцінювання, підвищує внутрішньорейтерну надійність за рахунок зменшення кількості випадкових помилок окремого рейтера та знижує рівень особистих неточностей стосовно різних критеріїв ситуації (наприклад, завдання, шкала чи кандидат). Для тестів високого статусу, зокрема,"й мінімально прийнятні показники надійності мають бути високими" [15, с.519].

Тим не менш, низка дослідників піддають сумніву ефективність вишколу рейтерів, а відтак і доцільність такого навчання. Побутує думка, що намагання змусити рейтерів до узгодженості щодо того чи іншого оцінювання може призвести до нівелювання їхніх професійних знань та досвіду, а відповідно і валідності їхніх суджень (зокрема, в оцінюванні письма) (Л.Барріт) [14, с.38]. Водночас С.Вейгл заперечує вагомість цього фактора для проведення вишколів, пояснюючи це тим, що досягнення згоди не є першочерговим завданням рейтера. Інше твердження про те, що у випадку оцінювання письма навчальні програми для рейтерів відволікають увагу останніх від змісту та загальної комунікативної ефективності зразків і змушують їх зосереджуватися на поверхневих аспектах (Д.Чарні) [14, с.38], визначених у дескрипторах шкали оцінювання, С.Вейгл також спростовує, зазначаючи, що таке спрощення є передумовою будь-якого вимірювання, а простота визначення необхідна для стандартизації використання шкали різними рейтерами.

Загалом вважається, що курси для рейтерів підвищують рівень систематичності їхньої поведінки. Так, хоч і незначна на сьогодні, але все ж певна кількість дослідників спробували оцінити довготривалість позитивного ефекту навчання та зробили невтішні висновки. Наприклад, М.Лунц і Дж.Шталь [14, с.39] виявили непостійність поведінки рейтерів вже за півдня після закінчення тренувань. Спостерігаючи за поведінкою екзаменаторів після вишколу з оцінювання письма, Т.Лумні і Т.Макнамара [14, с.39] помітили суттєві розбіжності між оцінками рейтерів під час симулятивного оцінювання протягом навчальної сесії та під час робочого оцінювання місяць потому. На їхню думку, покращити ситуацію могли б регулярні навчання рейтерів безпосередньо перед проведенням оцінювання та запровадження багаторейтерського оцінювання того самого зразка.

П.Конгдон і Дж.Макквін [14, с.39] аналогічно висловили занепокоєння відносними змінами у строгості рейтера при порівнянні одного дня навчання з наступним та їх абсолютними змінами при порівнянні першого дня курсу з останнім, що, на думку вчених, свідчить про недоцільність акредитації рейтерів після одноразового калібрування та потребу їх постійних вишколів, особливо при оцінюванні тестів з високим статусом. На підтвердження малоефективності навчальних курсів для рейтерів письма Дж.Кейсон наводить цифру у 35% випадків неточного оцінювання рейтерами навіть після проходження спеціального вишколу. А це спонукає до залучення до процесу оцінювання двох чи більше незалежних рейтерів для забезпечення надійності результату. У випадку масштабних тестів з високим статусом С.Елдер радить також застосовувати статистичні програми на зразок FACETS та метричної системи Джорджа Раша, які дозволяють корегувати різницю між індивідуальними оцінюваннями різних рейтерів, враховуючи особливості їхньої роботи [14, с.39]. Щоправда, такі корективи можливі лише у випадку особистої послідовності у строгості/поблажливості рейтерів.

Викладений матеріал дозволяє зробити наступні **висновки.** Попри виклики, пов'язані з проведенням тестування – інструменти (вибір та впровадження), середовище (програмне та матеріальне забезпечення), час (обмеженість в часі для розробки, проведення та аналізу) – воно залишається найпоширенішим засобом оцінювання іншомовної компетентності. У випадку масштабних тестів з високим статусом зазначені фактори мають особливе значення. Ефективність такого оцінювання безпосередньо залежить не лише від валідності та надійності самого тесту, а й від валідності та надійності його оцінювання. Відтак попередній аналіз проведених досліджень довів, що професійна підготовка рейтерів може суттєво підвищити ці показники, рівно ж як і внутрішньорейтерську послідовність та міжрейтерську надійність. Нечисельні думки щодо малоефективності таких навчань чи їх короткотривалої ефективності потребують подальшого емпіричного вивчення і підтвердження. Проте попередньо погоджуємося, що такі недоліки можна спробувати мінімізувати за рахунок полірейтерського оцінювання та систематичній рестандартизації самих рейтерів, особливо безпосередньо перед оцінюванням тестів з високим статусом.

## Список використаної літератури


1. Прісовська Г.Є. Типологія тестового оцінювання / Г.Є.Прісовська // Новітні тенденції навчання іноземної мови за професійним спрямуванням: Матеріали всеукраїнської науково-практичної конференції. – 2011. – С.88-89
2. Копилова І.М. Тестування умінь рецептивних видів мовленнєвої діяльності студентів / І.М.Копилова [Електронний ресурс]. – URL http://intkonf.org/kopilova-im-testuvannya-umin-retseptivnih-vidiv-movlennevoyi-diyalnosti-studentiv
3. Гарматюк Н.Д. Особливості застосування тестового контролю при вивченні іноземної мови у вищих навчальних закладах / Н.Д.Гарматюк, В.П.Марценюк. // Медична освіта. – 2013. – №3. – С.17–24
4. Кладкова Т.А. Комплексний контроль і засоби його проведення у процесі вивчення іноземних мов майбутніми психологами / Т.А.Кладкова [Електронний ресурс]. – URL: www.seanewdim.com
5. Knoch U. An evaluation of an online rater training program for the speaking and writing sub-tests of the Aptis test./ U. Knoch, J. Fairbairn & A. Huisman // Papers in Language Testing and Assessment. – 2016. – Vol. 5, Issue 1. – PP.90-106
6. Бобошко Т.М. Логічний підхід до аналізу категорії оцінки / Т.М.Бобошко. // Іноземні мови у вищому навчальному закладі: теоретичні засади та прикладні аспекти: Матеріали Всеукраїнської науково-теоретичної конференції. – 2013. – С.8–10
7. Fahim M. The Effects of Rater Training on Raters' Severity and Bias in Second Language Writing Assessment / M.Fahim, H.Bijani // Iranian Journal of Language Testing. – 2011. – Vol. 1, No. 1. – PP.1-16
8. Knoch U. An Evaluation of the Effectiveness of Training Aptis Raters Online. Final report / U.Knoch, J.Fairbairn, A.Huisman // Language Testing Research Centre. The University of Melbourne. British Council. – 2015. – 21p.
9. Pufpaff L. A. The Effects of Rater Training on Inter-Rater Agreement / L.A.Pufpaff, L.Clarke, R.E.Jones // Mid-Western Educational Researcher. – 2015. – Volume 27, Issue 2 117. – PP.117-141
10. Грищук Ю.В. Забезпечення валідності та надійності мовних тестів згідно вимог стандартів НАТО «STANAG 6001» / Ю.В.Грищук, В.Д.Галкіна // Вісник НТУУ "КПІ". Філософія. Психологія. Педагогіка. – 2015. – №1. – С.111-114
11. Kondo Y. Examination of rater training effect and rater eligibility in L2 performance assessment / Y.Kondo // Journal of Pan-Pacific Association of Applied Linguistics. – 2010. – №14(2). – PP.1-23
12. Пасічник М.В. Розробка ефективних тестів для контролю читання англомовних професійно орієнтованих текстів / М.В.Пасічник, С.Е.Жигалко // Молодий вчений. – 2015. – № 7 (22) Частина 2. – С.66-69
13. Csépes I. Measuring Oral Proficiency Through Paired-task Performance Language Testing and Evaluation (Book 14) / I.Csépes // Peter Lang GmbH. – 2009. – 240p.
14. Elder C. Evaluating rater responses to anonline training program for L2writing assessment / C.Elder, G.Barkhuizen, U.Knoch, J.von Randow // Language Testing. – January 2007. – №24 (1). – PP.37–64
15. Проценко Н.В. Тестовий контроль та сучасні технології викладання іноземних мов / Н.В.Проценко [Електронний ресурс]. – URL: http://nbuv.gov.ua/UJRN/Nznuoaf_2009_11_90


Рецензент: докт.пед.наук, проф. Микитенко Н.О.

Стаття надійшла до редакції 19.10.2017 р.
Стаття прийнята до друку 24.10.2017 р.

**Бурак Марьяна**
аспирант кафедры английской филологии
Тернопольский национальный педагогический университет имени В.Гнатюка г.Тернополь, Украина

## СПОСОБЫ МИНИМИЗАЦИИ ВЛИЯНИЯ РЕЙТЕРОВ НА ВАЛИДНОСТЬ И НАДЕЖНОСТЬ ОЦЕНИВАНИЯ ЯЗЫКОВЫХ МАСШТАБНЫХ СТАНДАРТИЗИРОВАННЫХ ТЕСТОВ

*Статья рассматривает основные факторы влияния рейтеров на валидность и надежность результатов субъективного оценивания продуктивных видов иноязычной коммуникативной компетентности, рассматривает разные аспекты тестирования как самого распространенного средства определения уровня владения языком и анализирует исследования эффективности учебных программ для рейтеров для*

*минимизации рейтерской неточности и непоследованости в оценивании, что особенно важно в случаи масштабных стандартизированных тестов.*

*Ключевые слова: тестирование, валидность, надежность, последовательность, рейтер, иноязычная компетентность, оценивание*

**Burak Mariana**
Post-Graduate Student
Ternopil Volodymyr Hnatyuk National Pedagogical University, Ternopil, Ukraine

## WAYS TO MINIMIZE RATERS' EFFECTS ON THE VALIDITY AND RELIABILITY OF LANGUAGE STANDARDIZED TEST ASSESSMENT

*The article highlights the main rater effects on the validity and reliability of foreign language proficiency subjective scoring, describes different aspects of testing as one of the most wide-spread means of language assessment and provides analysis of the existing researches on the efficiency of rater training programs in terms of minimizing raters' biases and consistency in the process of rating. The objectivity of the proficiency evaluation is of special importance in case of high-stakes examinations such as TOEFL, TOEIC, IELTS, Cambridge English Test, CaMLA, OPI, which demand extreme accuracy of their results. Training raters seems to be the most important way of achieving reliability of the assessment. However, there are few studies researching the exact affects of such sessions as well as the degree of their effectiveness in terms of reducing the raters' biasedness.*

*Key words: testing, validity, reliability, consistency, rater, L2 proficiency, assessment*