

МІНІСТЕРСТВО ОСВІТИ І НАУКИ, МОЛОДІ ТА СПОРТУ
УКРАЇНИ

Державний вищий навчальний заклад
«ПРИКАРПАТСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАСИЛЯ СТЕФАНІКА»

КАФЕДРА АНГЛІЙСЬКОЇ ФІЛОЛОГІЇ

**ВСТУП ДО МОВНОГО ТЕСТУВАННЯ
ТА ОЦІНЮВАННЯ**

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ ДО КУРСУ

Івано-Франківськ – 2013

УДК 371.214.114:378.14:811.111

ББК 74.580.263

В 85

Вступ до мовного тестування та оцінювання: методичні рекомендації до курсу / [упоряд. І. М. Романишин]. – Івано-Франківськ : НАІР, 2013. – 56 с.

Дане видання містить рекомендації щодо організації вивчення курсу "Вступ до мовного тестування та оцінювання" студентами п'ятого курсу спеціальності 7.0203030204. Філологія «Мова та література (англійська)». У ньому розкриваються мета і завдання дисципліни, подано назви тем і перелік питань лекційних і практичних занять і тих, що виносяться на самостійне вивчення, рекомендовану літературу і матеріали довідкового характеру, що, в цілому, сприятимуть ефективності викладання предмету та підвищенню рівня професійної підготовки студентів.

Рецензенти:

Українська О.О. – кандидат педагогічних наук, доцент кафедри англійської філології Харківського національного педагогічного університету імені Григорія Сковороди

Яців С.О. – кандидат педагогічних наук, доцент кафедри англійської філології Прикарпатського національного університету імені Василя Стефаника

Друкується за ухвалою вченої ради факультету іноземних мов Державного вищого навчального закладу «Прикарпатський національний університет імені Василя Стефаника»

Івано-Франківськ, 2013

© Романишин І. М.

ВСТУП

Дисципліна «Вступ до мовного тестування та оцінювання» викладається на п'ятому році навчання для студентів спеціальності 7.0203030204. Філологія за спеціальністю «Мова та література (англійська)».

Основною метою курсу є ознайомити студентів з теоретичними основами тестування та оцінювання, європейськими стандартами в галузі моніторингу мовної освіти. Допомогти студентам розвинути критичне розуміння основних принципів і загальноприйнятої практики мовного тестування та оцінювання. Сформувати в студентів практичні навички та вміння в складанні тестових завдань та оцінюванні якості їх змісту.

Завдання курсу – забезпечити студентів необхідними знаннями з предмету вивчення та сформувати в них практичні вміння, які вони зможуть застосувати в професійній діяльності вчителя/викладача англійської мови та літератури.

У результаті вивчення навчальної дисципліни студент повинен

знати:

Мету та завдання курсу

Основні положення та принципи тестового контролю знань

Методи тестування

Типи оцінювання

Компоненти тестових завдань різного формату

Критерії оцінювання завдань відкритого типу

Лінгвістичні основи оцінювання рівня сформованості іншомовної комунікативної компетенції

Вміння з читання, аудіювання, говоріння, письма та загального використання мови, що оцінюються тестовим методом

Основні типи тестових завдань

Технологію складання тестових завдань

Вимоги щодо укладання та дизайну тестових завдань основних типів.

Основні методи статистичного аналізу тестових завдань

Базову термінологію з галузі мовного тестування та оцінювання

Європейські стандарти в галузі мовної освіти

Українське законодавство, що регламентує незалежне оцінювання якості знань учнів загальноосвітніх навчальних закладів з іноземних мов

вміти:

Узагальнювати отримані теоретичні знання
Застосувати на практиці теоретичні положення стандартизованого контролю знань
Вести пошук проблемних питань, що винесені на самостійне чи індивідуальне опрацювання
Визначати рівень складності текстів з читання та аудіювання відповідно до шкали Загальноєвропейських мовних рекомендацій
Оцінювати завдання відкритого типу відповідно до критеріїв
Складати тестові завдання різних форматів з основних видів мовленнєвої діяльності та загального використання мови
Аналізувати тестові завдання укладені колегами з рекомендаціями щодо покращення їх якості
Визначати рівень складності та надійності тестів
Презентувати власні тестові завдання
Давати критичну оцінку опублікованим тестовим завданням, які рекомендуються для використання у навчальному процесі

Курс читається англійською мовою, що сприяє збагаченню словникового запасу студентів відповідною термінологією та позитивно впливатиме на оволодіння ними усним і писемним англійським мовленням.

Лекційні заняття (16 годин) покликані підвищити інтерес студентів до глибшого самостійного вивчення питань курсу, сприятимуть формуванню їхньої професійної компетенції.

Практичні заняття (8 годин) закріплять набуті на лекціях і під час самостійної підготовки знання з курсу, покращать професійно-орієнтовані комунікативні та інформаційно-комунікаційні вміння студентів.

Курс «Вступ до мовного тестування та оцінювання», розроблений автором, відноситься до дисциплін за вибором студента. Він читається в десятому навчальному семестрі в обсязі 54 години і закінчується теоретичним і практичним модулем. Вивченню дисципліни передують базові для неї курси педагогіки

(3-й семестр), методики викладання основної іноземної мови в загальноосвітніх навчальних закладах (7-й семестр), інші предмети лінгвістичного циклу (5-9 семестри).

Лекційна тематика базується на принципових положеннях лінгводидактики і має науково-пізнавальну спрямованість. Практичні заняття покликані сформувати в майбутніх фахівців практичні навички та вміння складати та презентувати тестові завдання різних форматів, оцінювати якість змісту завдань, укладених однокласниками. Самостійна робота в обсязі 30 годин покликана допомогти студентам узагальнити теоретичні знання та закріпити набуті практичні навички та вміння з відбору текстів для лінгводидактичного тесту, написання тестових завдань, аналізу їх якості, оцінювання завдань відкритого типу за критеріями тощо.

Система **контролю знань** студентів здійснюється через:

- поточний контроль – усне опитування на лекціях і практичних заняттях через індивідуальні, парні та групові форми роботи; тестові завдання, які мають на меті перевірити рівень теоретичних знань з курсу;
- виконання індивідуальних завдань, які покликані перевірити навички та вміння застосовувати на практиці принципи стандартизованого контролю знань;
- підсумковий контроль – один змістовий модуль упродовж семестру.

Методичні рекомендації складаються зі вступу, двох основних частин, списку рекомендованої літератури до всього курсу, додатків.

У вступі розглядаються мета і завдання курсу, мета лекційних і практичних занять, перелік знань і вмінь, які повинні бути/сформувати студенти в результаті вивчення дисципліни, види контролю набутих знань і вмінь студентів. У першій частині дається тематика і перелік питань, які розглядаються в аудиторії під час лекційних практичних занять. Друга частина містить матеріали довідкового характеру: короткий глосарій, рекомендації щодо написання тестових завдань з читання, аудіювання, загального використання мови, усного та писемного мовлення, формату множинного вибору, інформацію про види статистичного аналізу якісних показників тесту з коментарем і прикладами щодо їх застосування. Додатки містять таблиці, у яких зведені тематика і

перелік питань підсумкового змістового модулю, а також зміст, обсяг і форми контролю самостійної навчальної роботи студентів.

Дане методичне видання покликане сприяти ефективності викладання курсу «Вступ до мовного тестування та оцінювання», розширенню кругозору та підвищенню професійного рівня підготовки студентів спеціальності «Мова та література (англійська)».

ЧАСТИНА 1
ТЕМАТИКА ЛЕКЦІЙ

LECTURE 1

Subject and Aims of the Course “Introduction to Language Testing and Assessment”

Points for discussion:

1. Importance of the course
2. The purpose and aims of the discipline; its relation to other subjects
3. European standards in language testing: the CEF
 - 3.1 Background to the CEF
 - 3.2 The aims of language learning as stated in the CEF
 - 3.3 The CEF and learner’s language competences
 - 3.4 The CEF levels of proficiency in a foreign language
4. Ukrainian School-leaving English Exam

LECTURE 2

Basic Testing Terminology

Points for discussion:

1. Text selection and task design terminology
2. Task formats terminology
3. The language of data collection

LECTURE 3

Insight into Assessment

Points for discussion:

1. Key issues: Evaluation, Assessment, Testing (definitions)
2. Assessment: Purposes and focuses
 - 2.1 Fundamental concepts of assessment
 - 2.2 Other concepts of assessment
 - 2.3 Types of assessment
3. Language testing
 - 3.1 What is a test? A language test?
 - 3.2 Types of tests (in terms of method and purpose)
 - 3.3 Task formats/Item types

LECTURE 4

Test Specifications & Test Construction Process

Points for discussion:

1. Test specifications
 - 1.1 Who needs them?
 - 1.2 What should a TS include?
2. Test construction process

LECTURE 5

Testing Reading

Points for discussion:

1. What is reading?

2. Purposes of reading tests
3. Texts for Reading Paper
4. Reading skills and strategies
5. Levels of the reading competence
6. Guidelines for Reading Paper
7. Types of reading tasks
8. Standardized rubrics
9. Task components
10. Requirements for Multiple-Choice item-writing

LECTURE 6

Testing Language in Use

Points for discussion:

1. What is testing language in use?
2. The difference between a Reading test and a Use of English paper
3. What is tested in Use of English exams?
4. Task/question types (open gap-filling, banked gap-filling, multiple choice, spot the error, word formation)
5. Guidelines for Use of English Paper

LECTURE 7

Item Writing - Best Practices

Points for discussion:

1. General Guidelines for Item Writing as Derived From The Big Five
2. Ten Commandments for Item Writers

3. Feedback

LECTURE 8

Test Analysis

Points for discussion:

1. Methods of data collection

1.1 Questionnaires

1.2 Think aloud protocols

2. Methods of test analysis

2.1 Facility Value

2.2 Discrimination Index

2.3 Distribution statistics: the mean, the mode, the median, the range, the deviation index

3. Washback and impact

ТЕМАТИКА ПРАКТИЧНИХ ЗАНЯТЬ

WORKSHOP 1

Points for consideration:

1. Importance of the course.

2. The purpose and aims of the discipline; its relation to other subjects.

3. The aims of language learning as stated in the CEF.

4. The CEF and learner's language competences.

5. The CEF levels of proficiency in a foreign language.

6. Basic Testing Terminology.

7. Types of tests.

8. Fundamental concepts of assessment (Validity. Reliability. Feasibility).
9. Types of assessment.

WORKSHOP 2

Points for consideration:

1. Text level selection for reading and listening tasks.
2. Looking for test errors.

WORKSHOP 3

Points for consideration:

1. Reading (listening) tasks design.
2. Use of English tasks design.

WORKSHOP 4

Points for consideration:

1. Writing tasks design.
2. Speaking tasks design.

Рекомендована література

Базова

1. Гронлунд, Норман Е. Оцінювання студентської успішності: Практик. посіб. – К.: Навчально-методичний центр «Консорціум із удосконалення менеджмент-освіти в Україні», 2005. – 312 с.

2. Квасова О. Г. Основи тестування іншомовних навичок і вмінь: Навчальний посібник / О. Г. Квасова. – К. : Ленвіт, 2009. – 119 с.
3. Alderson, C.J. 2005. *Assessing Reading*. New York: Cambridge University Press.
4. Alderson, J. C, Clapham, C, & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge University Press.
5. Bachman, L.F. 1990. *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.
6. Bachman, L.F. and Palmer, A.S. 1996. *Language Testing in Practice*. Oxford: Oxford University Press.
7. Fulcher, G. (2003). *Testing Second Language Speaking*. Longman.
8. Heaton, J. B. (1988). *Writing English Language Tests*. Longman.
9. McNamara, Tim. 2000. *Language Testing*. Oxford: Oxford University Press.
10. Luoma, S. (2004). *Assessing Speaking*. J. C. Alderson & L. F. Bachman (Series Eds.)
11. McNamara, T. F. (1996). *Measuring Second Language Performance*. Longman.
12. Weir, C. (1990). *Communicative Language Testing*. Prentice Hall International Ltd.

Допоміжна

1. Державний стандарт базової і повної загальної середньої освіти (Освітня галузь «Мова і література»: іноземні мови) // Іноземні мови в навчальних закладах. – 2004. - № 1. – С. 10-19.
2. Програми для загальноосвітніх навчальних закладів та спеціалізованих шкіл з поглибленим вивченням іноземних мов. – Ірпінь : Перун, 2005. – 208 с.
3. J.Charles Alderson and Maria Cseresznyes. 2000. *Into Europe. Prepare for Modern English Exams. Reading and Use of English*. Budapest: The Teleki Laszlo Foundation.
4. Spaan, Mary. 2006. *Test and Item Specifications Development*. In *Language Assessment Quarterly*, 3 (1), 71 –79.
5. Alderson, J. C, Nagy, E. & Oveges, E. (2000). *English Language Education in Hungary. Part 2. - Examining Hungarian Learners' Achievements in English*. Budapest: The British Council Hungary.
6. Brown, A. (2003). Interviewer variation and the co-construction of speaking proficiency. *Language Testing* 20, 1-25.

7. Csepes, I. (2002) Is testing speaking in pairs disadvantageous for students? A quantitative study of partner effects on oral test scores. *novELTy* 9(1), 22-45.
8. Egyud, Gy., & Glover, P. (2001). Oral testing in pairs - a secondary school perspective. *ELTJournal*, 55, 70-76.
9. Egyiid, G., Gal, I. & Glover, P. (Eds.) (2001). *English Language Education in Hungary. Part 3. - Training Teachers for New Examinations*. Budapest: The British Council Hungary.
10. Fekete, H., Major, E. & Nikolov, M. (Eds.) (1999). *English Language Education in Hungary. A Baseline Study*. Budapest: The British Council Hungary.
11. Fulcher, G. & Reiter, R. M. (2003). Task difficulty in speaking tests. *Language Testing* 20, 321-344.
12. Hughes, A. (1989). *Testing for Language Testers (2nd edition)*. Cambridge University Press.
13. Underhill, N. (1987). *Testing Spoken Language*. Cambridge University Press.

Інформаційні ресурси

Міністерство освіти і науки України: <http://www.mon.gov.ua/>

1. Український центр оцінювання якості освіти:
<http://www.testportal.org.ua/>

2. Британська Рада в Україні:
<http://www.britishcouncil.org/ukraine>

3. Проектні технології у викладанні ІМ:

<http://distant.ioso.ru/project/meth%20project/metod%20pro.htm>

4. З досвіду навчання лексики і граматики (англомовні матеріали):

<http://www.onestopenglish.com/section.asp?sectionType=listsummary&catid=58111>

<http://www.onestopenglish.com/section.asp?sectionType=listsummary&catid=58110>

<http://www.onestopenglish.com/section.asp?sectionType=listsummary&catid=58088>

<http://www.macmillandictionaries.com/resources/e-lessons/e-lessons.htm>

5. З досвіду навчання аудіювання, говоріння, читання, письма (англомовні матеріали):

<http://www.onestopenglish.com/section.asp?sectionType=listsummary&catid=59392>

<http://www.macmillandictionaries.com/resources/dictations/dictations.htm>

6. Автентичні аудіоматеріали для навчання АМ:

<http://www.onestopenglish.com/section.asp?sectionType=listsummary&catid=59782>

7. **Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, teaching, assessment.***

Cambridge: Cambridge University Press. [Електронний ресурс] :

http://www.coe.int/T/E/Cultural_Cooperation/education/Languages/Language_Policy/Common_Framework_of_Reference/1cadre.asp#TopOfPage

8. IELTS, TOEFL, Cambridge ESOL:

<http://www.onestopenglish.com/section.asp?sectionType=listsummary&catid=58065>

<http://www.onestopenglish.com/section.asp?sectionType=listsummary&catid=58063>

<http://www.ets.org/portal/site/ets/menuitem.fab2360b1645a1de9b3a0779f1751509/?vgnnextoid=69c0197a484f4010VgnVCM10000022f95190RCRD>

<http://www.ielts.org/>

<http://www.cambridgeesol.org/>

<http://www.ets.org/portal/site/ets/menuitem.fab2360b1645a1de9b3a0779f1751509/?vgnnextoid=06cfd898c84f4010VgnVCM10000022f95190RCR>

10. **Criteria for evaluating a speaking test**

<http://www.teachingenglish.org.uk/think/articles/evaluating-speaking-ielts-speaking-test>

11. **Reasons for testing and alternatives to testing**

<http://www.teachingenglish.org.uk/think/articles/testing-assessment>

ЧАСТИНА 2

ДОВІДКОВІ МАТЕРІАЛИ

GLOSSARY

Text Selection and Task Design Terminology

The glossary below is based on information in Chapter 4 of Bachmann and Palmer's *Language Testing in Practice* (OUP) 1996.

Accountability: the responsibility test designers have to candidates and stakeholders to be able to explain the rationale behind their assessment techniques.

Affective schemata (emotional correlates of topical knowledge) provide the basis on which candidates assess (consciously or unconsciously) the characteristics of a text/task and its setting in terms of past emotional experiences in similar contexts. Test designers must consider candidates' emotional responses to emotionally-charged topics.

Avoidance strategy: strategy adopted by learner to keep their linguistic output (spoken/written) within their communicative resources.

Band: a measure (e.g. 1 - 6 or A1 - C2) or description of the proficiency or ability of a test-taker, normally as described on some kind of scale and determined on the basis of test performance.

Bias: a factor or factors inherent within a test that systematically prevent access to valid estimates of candidates' ability.

Calibrate: to determine the value of test items against a particular measurement scale to reflect item difficulty.

Competence: the sum of knowledge, skills and characteristics that allow a person to perform actions.

Construct: those aspects of the candidate's underlying knowledge or skill which are the target of measurement in a test.

Criterion-referencing: using descriptions of what candidates should be able to do with the language in order to determine the pass score in a test.

Cut-off point: point on a scoring/scaling continuum at which a classification decision is made.

Descriptor: a statement which describes the level of performance required of candidates at each point on a proficiency scale.

Domain: the area of knowledge or skill or the set of tasks constituting criterion performance, and which is the target of the test.

Evaluation: the systematic gathering of information in order to make a decision.

Global assessment: the awarding of one or more scores to a piece of writing or spoken performance on the basis of an overall impression.

Item: an individual question in a test that requires the candidate to produce an answer.

Knowledge schemata (or topical knowledge) provides the information base which allows candidates to use language with reference to the world in which they live. Certain test tasks that presuppose cultural or topical knowledge may prove easier for candidates who have that knowledge and more difficult for any who don't.

Language ability is defined according to Bachman (1990) as involving two components language knowledge and strategic competence (i.e. a set of metacognitive strategies) which provide language users with the ability, or capacity, to create and interpret discourse in either test or non-test use.

Language knowledge is a domain of information in memory that is available for use by the metacognitive strategies in creating and interpreting discourse in language use. It includes the categories of organisational and pragmatic knowledge.

It is important for test designers to be fully aware of the full range of the components of language ability even though any given test may focus on only one or a few areas of language knowledge.

Mediation activity: activity where the language user acts as an intermediary between interlocutors who are unable to understand each other directly.

Norm-referencing: listing candidates in order of test results and passing them or failing them according to their position on the list.

Organizational knowledge involves a candidate's control of the formal structure of language for producing or comprehending grammatically acceptable utterances or sentences and for organising these to form both spoken and written texts. There are two areas of organizational knowledge: **grammatical** (vocabulary, syntax, phonology, and graphology) and **textual** which incorporates both the area of cohesion

and the knowledge of rhetorical (written) or conversational (spoken) organisation.

Plurilingualism: composite competence in the ability to use languages for the purposes of communication where a person viewed as a social agent has proficiency, to varying degrees, in several languages.

Practicality: all aspects concerning tests which concern time and resources.

Prompt: one or more sentences that provide candidates with a situation/topic on which to write, as well as directions on how to write it.

Reported score: the score which is reported to test candidates.

Sorting task: activity in which distinctions are made between descriptors to assign each to a particular level or category

Specifications: a document that sets out what a test is designed to measure and how it will be tested.

Strategic competence comprises a set of **metacognitive strategies** or higher order executive processes that provide a cognitive management function in language use and in other cognitive activities. Bachman and Palmer identify three general areas in which metacognitive strategies operate: goal setting, assessment, and planning.

Test difficulty: the extent to which a test or test item is within the ability range of a candidate or group of candidates.

Test discrimination: basic property of language tests in their attempt to capture the range of individual abilities.

Test handbook: a publication for stakeholders of a test (candidates, teachers) that contains information about the format and the content of a test.

Validation: process which involves gathering and evaluating the evidence for the reliability and validity of a test when used for a given purpose.

Task Format & Item Types Terminology

Cloze: task in which the candidate is expected to restore every nth word in a text.

Information transfer: task in which the candidate may be asked to

transfer information from a text to graphic form (chart, table, drawing, map etc.)

Limited/open-ended production response: another term for ‘short or restricted answer’ tasks.

Multiple choice: task format in which the candidate must choose the correct answer from a number of options only one of which is correct.

Multiple matching: an objective technique in which two sets of stimuli have to be matched against each other e.g. matching paragraph headings to their corresponding paragraphs.**Simulation:** a task that is very similar to a role-play in which candidates are required to act out an imagined communication situation.

Negative cloze: a task also known as ‘cloze elide’ or ‘intrusive word’ where candidates must delete words which the tester had added to a text, but which do not belong there.

Proof-reading: task format also known as ‘editing’ in which candidates must correct the mistakes in a text.

Sequencing: putting parts of a text into the correct order.

Summary: task in which candidates are asked to produce a precis or shortened version of a text or part of a written or spoken text.

Transformation: a grammar task in which candidates are asked to re-write or complete sentences so that the second sentence remains faithful in meaning to the original.

Word attack: task in which candidates are asked to work out the meanings of a selection of highlighted or boxed words within the context of a text.

Other Item Types/Tasks Formats:

C-test

Dialogue/Discourse completion

Dictation

Essay

Extra/intrusive word error detection

Gap-filling Open/Banked (word/sentence/paragraph)

Identify topic (text/paragraph)

Identify discrepancies

Listing
Matching banked words to definitions
Note expansion
Note-taking
Noticing Task
Phonemic discrimination
Problem solving
Punctuation
Register transfer
Response evaluation
Sentence transformation/Paraphrasing
Structure identification
Table completion
Translation
True/False
True/False/Not Given
Unique answer
Word associates
Word formation
Word sequencing (sentences)

Language of Data collection

Consistency: reliability or uniformity of successive responses.

Dependent variable: the principle attribute the researcher is investigating.

Dichotomous data: that which is either right or wrong.

Respondent: person who completes a questionnaire.

Instrument: a tool used to collect and organize information

Introspective: contemplation of one's own thoughts and feelings.

Likert: a type of response format used in surveys. Likert items have responses on a continuum and response categories such as "strongly agree," "agree," "disagree," and "strongly disagree."

Qualitative data: non-numerical data, e.g. open-ended questionnaire or interview data.

Quantitative data: numerical data obtained through counting and measurement, e.g. test scores and data from fixed response questionnaires.

Response rate: the proportion of subjects answering a questionnaire.

Tendency: an inclination or characteristic likelihood toward a certain condition or effect.

Think-aloud protocol: a method of eliciting data from subjects for research purposes, which involves subjects talking or thinking aloud during or shortly after completing an activity such as a test.

General Guidelines for Item Writing

Text selection

- All texts must be authentic.
- Texts must not be taken from printed material in other examinations or English language teaching materials.
- The topic of the texts must be accessible to the students' age group.
- Texts must not be offensive, distressing or violent.
- Texts must be of a suitable length.
- Texts must be of a suitable level of difficulty.
- A text must have a title unless it is used as part of the item.
- The same text can be used with more than one task. For example, a reading text with different tasks could be used for intermediate and more advanced levels, or a reading text used with intermediate students could be used as part of a Use of English task for more advanced students.

Task selection

- Have a clear idea what skill and sub-skill(s) you intend to test and select the task and text/prompts accordingly.
- The task type must be one which is familiar to the students who are to be tested.
- Tasks must be accessible to the students' age group.
- Tasks must not be offensive, distressing or violent.
- The time allocated for each task must be sufficient for somebody who has the ability being tested to complete the task comfortably within the time limit.
- Students must be able to see easily how the task relates to the text.
- No more than one testing technique should be used in each task. Do not switch from multiple-choice to single-word answers in a single task, for example.

Task design

- There should be a minimum of five items in a task.
- Items must follow the text sequence.
- Items must be spread evenly through the text.

- It must not be possible to answer any item without reference to the text. This must be checked carefully.
- In non-sequencing tasks, items must not be interdependent – students should not need one answer in order to find another.
- Items must not overlap.
- Two items must not have similar answers.
- Items must have a complete answer key. All possible good answers and predictable wrong answers should be provided.
- The items in each task must be numbered sequentially.
- One item in each task must have an answer provided as an example. The example item, written in italics, should be marked *0* and should precede the other items. If the options to be selected are marked with letters, the answer to the example should be marked with a letter according to its place among the options.
- The example answer must be provided in the form in which students are expected to write their answers and, where appropriate, it must be handwritten.
- “Find the wrong answer” type items are not acceptable in multiple-choice tasks.
- Each item must score one point.

Rubrics

- Rubrics should be in English. They must be clear, simple and brief. Redundancies must be deleted. Do not use exclamation marks and metalanguage.
- Conform to standard rubrics.
- Indicate clearly what students have to do i.e. what they have to write and where.
- The rubrics should be in bold.
- Put the total number of points available for each task (in other words, the number of items) at the end of the task, in bold.

Guidelines for Reading Paper

Text selection

- Avoid well-known novels as text sources.
- Do not use more illustrations than necessary. They should be clear and photocopiable.
- Do not use texts with content covered in detail in any textbook. You do not want to test knowledge of content, but ability to understand texts.
- Do not make any changes to the original text. Do not delete words, sentences or paragraphs from the selected body of text.
- If the text contains any offensive words or errors, only change these with great care and always seek the advice of a fellow teacher or a native speaker as to the acceptability of the changes you have made.
- If the text used is not the beginning part of a longer text, make sure it stands alone and does not contain references to previous parts of the original text.
- Edit the text layout so that it resembles the original text as closely as possible, but make sure this does not make the text too difficult to read (e.g. in gap filling tasks).

Task selection

- Have a clear idea about the purpose of each reading task, which reading skill is tested in a particular task, and why a particular task type is used.
- The task should reflect the way in which people usually read the type of text used.
- Consider applying more difficult texts with easier tasks and vice versa.

Task design

- Do not design in any one task more than 10 items for a passage of approximately 100 words.
- The wording of the items should be easy to understand, i.e. below the language level of the text, not require more reading than the text itself.
- Make sure there is only one possible answer to each question.
- Avoid questions which demand much imagination.

- Avoid Yes/No questions and True/False items, because guessing the right answer is easier.
- Do not delete the first words or beginnings of sentences.
- In any gap-filling task (e.g. modified cloze, banked cloze, discourse cloze) there should be a sentence at the beginning (before the example) for lead-in and a sentence at the end for lead-out, which do not have any items.
- In a gapped text, where you take out paragraphs, do not take out the first or the last paragraph, in order to leave enough context for comprehension.
- In gap-filling tasks there should be at least five words between gaps to provide enough context.
- In banked cloze tasks the items must be numbered in the gaps in brackets, and the options, marked with letters, must be listed in alphabetical order either at the bottom of the page or on the facing page.
- Provide four options for multiple-choice items. Multiple-choice questions should have only one correct answer and should not offer “either a) or b)” or “both a) and b)” type of answers as an option, because these are tricky.
- “Not mentioned” must not be used as an optional answer, because this can be misinterpreted. Many things “not mentioned” can still be inferred.
- In one task, the number of matches must not be more than 10 for intermediate students, and 12 for advanced students. Provide more choices than matches in matching tasks and banked gap-filling tasks (minimum 1, maximum 2 distracters to be provided).
- Make sure that ambiguous matches are excluded.
- Make sure that in sequencing tasks there is only one correct order and no more than 6-7 items for intermediate students, or 8-9 items for advanced students. Be careful if using newspaper articles in sequencing tasks: they are rarely in chronological order.
- Clearly indicate the required length of the answers (maximum 3 words) in open-ended (short-answer) questions.
- Design tasks with clear and consistent layout. The input text should retain the original format as much as possible. Paragraph texts clearly with indentations.

- Design one task (including rubric, text, questions, space or boxes for answers) to fit on one page. If this is not possible, the task must be arranged on two facing pages.

Rubrics

- Rubrics should be in English. They must be clear, simple and brief. Redundancies must be deleted. Do not use exclamation marks and metalanguage.
- Conform to standard rubrics.
- Indicate the type of text students are going to read and/or the text source (preferably both).
- Indicate clearly what students have to do, i.e. what they have to write and where.
- The rubrics should be in bold.
- Put the total number of points available for each task (in other words, the number of items) at the end of the task, in bold.

Guidelines for Use of English Paper

Text selection

- Choose relevant, interesting and intrinsically motivating topics for the age group of students concerned (for example, using computers, the internet, birds and animals, nature reserves, sports, tourism, traveling, archaeology, cuisine, education, books, language, and others).
- Texts for this kind of test may come from a variety of authentic sources. Texts taken from graded readers as they appear in the original are acceptable. Do not take texts or sentences from language teaching materials other than readers or from dictionaries.
- Use texts at an appropriate level of difficulty. In order to ensure that the task tests Use of English and not Reading, select texts the language level of which is below the tested level.
- Only make changes to the original text if absolutely essential. Do not delete words, sentences or paragraphs from the selected body of text. If the text contains any offensive words that should be replaced, or you feel that any other changes would be necessary, only make these changes with great care and always seek the advice of a fellow teacher or a native speaker as to the acceptability of the changes you have made.
- Do not use texts exceeding 300 words.
- Normally provide a title. It helps to contextualize the meaning.
- Edit the text layout so that it resembles the original text as closely as possible but make sure this does not make the text too difficult to read (e.g. in gap filling tasks).

Task selection

- The selected task type must be familiar to the students.
- Do not use “Find the wrong answers” multiple-choice tasks, because these are very tricky and may test intelligence rather than language.

Task design

- Test a range of different linguistic points in one task.
- Balance the number of lexical and structural items in text-based tasks (except in word transformation tasks).

- Use only one testing technique in one task.
- There should be a minimum of ten, a maximum of 20 items in a task, in an approximate ratio of one item per 10-15 words in text-based tasks.
- Highlight items clearly by using **boldface** (e.g. 1. He (**live**) here since he was born.).
- Provide an example in italics marked (0), and an example answer in a handwritten form. The example should follow the lead-in sentence.
- Make sure that each question can be answered independently; that is, the answer to one question does not depend on the answer to another question (except in sequencing tasks).
- Make sure that items do not overlap.
- Number items and use letters to mark options.
- Provide enough context in between gaps (at least five words) in gap-filling tasks.
- In any gap-filling task (e.g. modified cloze, banked cloze, discourse cloze) provide a sentence with no items at the beginning (before the example) for lead-in and one at the end for lead-out. This will help students understand the context of the items.
- In banked cloze tasks the items must be numbered in the gaps in brackets, and the options, marked with letters, must be listed in alphabetical order either at the bottom of the page or on the facing page.
- When designing multiple-choice cloze tasks, consult good monolingual dictionaries and books that give synonyms and near synonyms.
- Provide four options for multiple-choice items. Multiple-choice questions should have only one correct answer and should not offer “either a) or b)” or “both a) and b)” type of answers as an option, because these are tricky and difficult to understand.
- Put only one item in a line in intrusive-word tasks, “spot-the-error” and “spot-and-correct” tasks.
- Do not use an item at the beginning or at the end of a line in intrusive-word tests. Students should be required to write the word to be deleted in the space provided. Crossing out the intrusive word is not the right way to indicate the answer. Deleting an irrelevant word may make the student delete another word.

- In intrusive-word tests the word to be deleted must not appear in the line more than once.
- Design one task to fit on one page if possible, or on facing pages.
- Design tasks with clear and consistent layout. The input text should retain the original format as much as possible. Paragraph texts clearly with indentations.
- Provide enough space for answers.
- Provide answer keys reflecting the form in which answers are expected. Give all possible or acceptable answers. Unacceptable answers that are likely to occur should also be indicated.
- Each item scores one point.

Rubrics

- Give the number of errors in “spot-the-error” tasks, and indicate that students must tick error-free lines.
- In “spot-the-error” tasks identified errors and identified error-free lines each score 1 point.
- In “spot-and-correct” tasks only corrected items score; no marks are given for errors identified but not corrected.

Guidelines for Listening Paper

Text selection

- Use spoken texts that sound like spoken English. Where, for example, pauses or fillers would be normal in conversation, they should also be used in the recording. You may record the listening texts from radio stations off-air, or from audio files available on the Internet. You may also record your own texts, using a good microphone and a digital recorder.
- Avoid using texts from television programmes. These texts have been recorded for TV viewers who listen and watch at the same time. Therefore, these texts consist of two sources of information, one is the picture and the other is the sound. If one is missing, comprehension becomes very difficult or even impossible.
- Select texts whose topic is accessible to the students' age group.
- Try to select texts that are likely to be interesting for students. Be aware that humorous texts, for example radio comedy shows, might be enjoyable and easy for you but very difficult for students.
- Avoid offensive and distressing topics, such as war, fatal diseases, family tragedies, etc.
- Avoid well-known audio recorded novels or any other well-known radio programmes as text sources.
- Avoid listening materials that were recorded and published for teaching.
- Select texts of suitable length. If the text is a part of a longer text, make sure it stands alone and does not contain references to previous or following parts of the original texts.
- Do not use more illustrations than necessary. When illustrations are used, they should be clear and photocopiable.
- Do not use texts with content covered in detail in any listening exercise included in a language textbook. You do not want to test knowledge of content, but ability to understand recorded texts.
- Select texts that do not assume a high level of background knowledge (social, cultural, factual etc.).
- If you decide to produce a text of your own, you must seek the advice of a fellow teacher and, if at all possible, of a native speaker as to the accuracy and the acceptability of the vocabulary and grammatical structures used in the spoken texts.

- Ensure that in case of sequencing tasks the text lends itself to sequencing.
- Do not use listening texts with many speakers participating in a conversation, especially if it is not absolutely clear who says what, or where the relationships are not explicitly stated.
- Ideally each speaker's voice should be clearly distinguished from the others. For example, a male could be paired with a female, or a younger with an older person.

Task selection

- Have a clear idea about the purpose of each listening task, which listening skill is to be tested in a particular task (e.g. listening for the main idea or listening for specific information), and why a particular task type is used.
- The task should reflect the way in which people usually listen to the type of text used.
- The students should be familiarised with the task type in advance.
- Tasks should not test students' memory.
- Consider applying more difficult texts with easier tasks and vice versa.

Task design

- Produce tasks from tapes, not tape-scripts.
- There should be a minimum of five items in a task.
- The wording of the items should be such that it is easy to understand and does not require too much reading.
- Make sure items cannot be answered correctly without listening to the text.
- Use clear signposts to help students in tasks where texts consist of a number of segments.
- Provide an example, written in italics and marked with (0), at the beginning of each task. Example answers should be given in a hand-written form where appropriate.
- Allow enough time for students to familiarise themselves with items before listening to the text. Intermediate: 4 seconds per item; advanced: 3 seconds per item.

- Allow enough time for students to finalise their answers after the text is finished: 3 seconds per item.
- Make sure that each question can be answered independently; that is, the answer to one question does not depend on the answer to another question (except in sequencing tasks).
- Select items in such a way that there is enough time for students to answer before having to concentrate on the text again for the next item.
- Spread items evenly. Do not have long stretches of text where no item needs to be answered.
- Sequence items following the sequence of ideas in the text.
- Number items and use letters to mark options.
- Make sure there is only one possible answer to each question. If more than one answer is required, this fact must be clearly indicated, and they must count as different items.
- Do not focus only on one type of factual information (numerals, countries, nationalities etc.).
- Do not base items on proper names and avoid items the answer to which depends on understanding proper names (e.g. people's names, geographical names, etc.).
- Provide four options for multiple choice items but make sure they do not require too much reading.
- Provide more options than pairs in matching tasks.
- Make sure you base completion items on factual information, and not on subjective judgment. The missing words should be concrete nouns and verbs rather than adjectives because adjectives usually have synonyms and many other plausible replacements.
- In sentence completion tasks the sentences candidates hear should not be the same as the sentences they have to complete.
- In productive tasks (i.e. in completion tasks and short-answer question tasks) avoid items whose answer is very difficult to spell.
- Design one task to fit on one page or two facing pages.
- Design tasks with clear and consistent layout.
- Provide answer keys reflecting the form in which answers are expected. Give all possible or acceptable answers. Unacceptable answers that are likely to occur should also be indicated. In tasks where the correct answer may be phrased in various ways, the

answer key should provide content-related instructions concerning acceptable answers.

- Weigh items equally (1 item = 1 point).

Rubrics

- Use simple, standardized rubrics.
- Word all instructions in English.
- Present all instructions, identical with the taped rubrics, on the test sheet.
- Inform candidates what they are going to hear: the text type and /or source.
- Inform candidates in the rubrics when they are expected to answer, and how many times the text will be played.
- Indicate clearly what candidates are expected to do i.e. what they have to write and where.
- Give guidance on the length of appropriate answers.
- Rubrics must be written in bold.
- Indicate the total number of points available for each task at the end of the task in bold.

Checklist

Text

- Is the text type worth testing?
- Is the text user-friendly in style and acceptable in content?
- Is the length of text appropriate for your students?
- Is the level of difficulty appropriate?
- Do any alterations to the text conform to the Guidelines?

Task

- Is the rubric clear, brief and in bold?
- Has an example been provided for each task?
- Is there a complete answer key for each task?
- Is the task type familiar to students?
- Is the task accessible and acceptable?
- Is the number of items acceptable?
- Are the items spread evenly through the text?

- Are all items properly labeled: questions with numbers, options with letters?
- Can each question be answered independently?
- Do items follow the text sequence?
- Do answer keys reflect the form in which the answer is expected?
- Does the answer key include alternative answers (if applicable)?

Requirements for multiple choice item writing

Stem

- Question has been clearly defined in the item stem.
- Include as much of the item in the stem as possible.
- Do not leave a blank for completion at the beginning of the stem.
- Omit any irrelevant material in the item stem.
- Avoid grammatical cues in the item stem.
- Minimise the number of negatively stated items.
- If using a negative in the stem, emphasize it in capitalized bold typeface (e.g. NOT).

Options

- Avoid correct options which are noticeably longer or shorter than the distractors. Avoid stating the correct option in more detail than the incorrect options.
- Avoid generalising the correct option so that it has wider application than the distractors.
- Avoid correct options that contain familiar or stereotyped phraseology. Ensure all options are grammatically consistent with the stem.
- Vary the position of the key for each item.
- List options on separate lines, beneath each other.
- Avoid incorrect options which contain language or technical terms with which candidate may be unfamiliar.
- Avoid distractors which are clearly unreasonable statements.
- Do not repeat key word(s)/key phrase from the stem in the correct option.
- Avoid using specific determiners like *sometimes* and *never* in the options.
- Avoid the options all of the above and none of the above.
- Avoid slang correct options.
- If the item is testing the definition of a word, put the word to be defined in the stem and make the options alternative definitions or meanings.

- Write at least three distractors for every question.
- Make each distractor plausible and attractive to candidates.
- When possible, arrange options in a logical and sequential order.
- Ensure all options are independent of each other.
- Ensure there is only ONE correct answer.
- Ensure you're testing comprehension of the text and not of the items.
- Do not use any trick items.
- Ensure none of the options is ambiguous.
- Pre-test all tasks.

Guidelines for Writing Paper

Prompt Selection

- Use prompts that are clear. Consider using pictures / drawings as prompts, but ensure that they are black and white, photocopyable (in size max. A4) and appropriate for the age.
- Provide enough information in the input but not so much that the task becomes a reading task.
- Where possible, input texts should be authentic (e.g. advertisements, articles, leaflets, brochures, letters, etc.). Where this is not possible (e.g. personal notes, diary entries, etc.) make sure the text is written in correct and appropriate English.
- The language level of verbal prompts should be below the tested level.
- The input text (edited/simplified versions of authentic texts) is as short as possible. Ideally, it should not be longer than 50 words.

Task Selection

- The task should be realistic and plausible, and should be within the candidate's range of experience.
- Do not test the candidate's background knowledge of the world.
- Make sure that the task does not require creativity or imagination from students. You should provide appropriate prompts which determine what students should write, give enough information for them to write about, and can elicit the required amount of language.
- Make sure that if personal data or information is required, this should be provided in an input text or in other prompt(s) so as to standardise correct answers.
- Special care must be taken that the task does not turn into copying.

Task Design

- In order to make the task realistic and communicative, create a complete task environment: it should be clear for students to whom, why, when, where, about what and how they are supposed to write.
- The required length must conform to the Specifications.
- Edit each task on one A4 page, or two facing pages.
- Provide enough space to complete tasks. The space where students respond must be lined.

- If there is an authentic input text, edit its layout so that it resembles the original as closely as possible.

Rubrics

- Use standardised and simplified rubrics, well under the candidate's expected level.
- Avoid using exclamation marks and repetitions.
- Give all rubrics in English.
- Indicate the reason for writing, the audience and the required text type.
- Word rubrics so that they cannot be copied to complete the task.
- If students are required to write about several things, organise these requirements in bullet points (rather than list them in a continuous line) so that they are easier to follow.
- Indicate the required length in number of words.
- Indicate where to write answers.

Guidelines for Speaking Test

Speaking tasks should

- measure the intended proficiency level and skills;
- provide the widest scope for candidates to perform;
- provide candidates with an opportunity to contribute equally to the exchange (in the paired format);
- be realistic in the sense that they should replicate real life language use as closely as possible;
- explore themes that candidates can relate to and/or on which they can express an opinion;
- be fair in the sense that they should not be biased towards a specific group of candidates;
- generate a rateable language sample from candidates.

Guidelines for Writing Paired Tasks

Item writers should bear in mind that

- the reading input given on the candidate's task sheet should be minimal as the exam is not intended to measure candidates' reading ability (use word prompts instead of full sentences, maximum 30 words);
- the task instructions should be read out by the Interlocutor and so they cannot be given on the Task Sheet for the candidates (they should appear on the Interlocutor's Sheet only);
- candidates' potential contributions to the interaction should be
- balanced, which could be achieved by giving an equal number of visual or word prompts for both candidates;
- candidates should have comparable tasks, i.e. both of them should be required to do the same thing (to list, compare, contrast, select, justify, modify, extend, reduce options, etc.) in order to facilitate a balanced, realistic and smoothly-running exchange between them;
- the interaction between the candidates should be task-based as it seems to give them a meaningful purpose to engage in a conversation (e.g. select the three most/least important aspects of something, rank order items in a given list);

- the tasks have to be guided but not fully controlled, i.e. candidates should have a chance to add something of their own to the exchange (use the word prompt Anything else? to indicate that candidates can add their own ideas);
- controversial topics seem to be more likely to generate a discussion than neutral topics;
- distressing, offensive, violent or taboo topics should be avoided;
- the topic should be something on which it is reasonable to expect candidates to have an opinion;
- in order to make the task realistic for candidates, they should define the context of the conversation very carefully, but at the same time they should not force candidates to agree with a point of view that they may not accept in real life. Candidates should be given a chance to voice their own opinion rather than argue for something that they cannot identify with.

General Guidelines for Designing Role-play Tasks

Prompts

- Select prompts/input text and design the task so that it generates the appropriate amount of varied language at the required level.
- Use prompts that are clear, black and white, photocopyable (in size max. A4) and appropriate for the target age.
- Limit input text so that candidates can process it in 30 seconds.
- The language level of verbal prompts should be below the tested level.
- Do not use distressing, offensive, violent or taboo topics. Treat sensitive topics with care.
- Do not use surreal, abstract, puzzling or symbolic pictures at lower levels.

Situations, roles, interaction

- Create situations that are life-like, suitable for the target age-group and in which it is realistic for them to speak in English.

- Candidates should not be required to take unfamiliar roles.
- Use different questions and tasks of different complexity if using the same prompt for more than one level.
- The task should not require students to use their imagination or creativity as these are not to be tested.
- Role-plays should involve real interaction and have an outcome.
- Create an information-gap or opinion-gap in role-play tasks.
- In the individual format, when the student has to speak to the interlocutor, make sure the interlocutor does not have to speak more than the student.
- In the paired format, when two candidates have to talk to each other, make sure they have equal roles.
- Restrict the use of yes/no and alternative questions.
- Always try out the task yourself to see whether you can speak for the required length of time.

Rubrics

- Use standardised instructions.
- Instructions for candidates should be given orally by the interlocutor they should appear only on the interlocutor's sheet. Short, simplified versions only may appear on the candidate's sheet.
- Instructions must be given in English.
- Instructions should be simple.

CEFR Descriptors

Feedback Grid

Task ID Number _____

Level Targeted _____

Text Title _____

Task Components	Layout	Content
Rubric		
Text		
Items		
Example		
Answer Grid		
Key		
Total Score		

Any other comments:

The feedback given by (Student's name, Group Index) _____

Methods of Test Analysis

The facility value (F.V.)

- measures the level of difficulty of an item
- An item's facility value is the percentage of students to answer it correctly.

(E.g. If there are 300 students and 150 of them get the item right, the F.V. of the item is $150/300$, which is 50% (often shown as a proportion: .5).

Possible variations of the F.V.

Intervals of Difficulty Index	Test Characteristics
Over 80	Very easy
60-79	Easy
40-59	Optimal
20-39	Difficult
19 and lower	Very difficult

Discrimination Index (D.I.)

- measures the extent to which the results of an individual item correlate with results from the whole test.
- discriminates between students at different levels of ability.

(If the strongest students get an item wrong, while the weakest students get it right, there is clearly a problem with the item, and it needs investigating.)

Ways of calculating a discrimination index

1. Rank the students according to their total score.
2. Divide them into three groups, making sure that the top and bottom groups have equal numbers of students.
3. Count how many students in the top group get an item right, and how many in the bottom group.

4. Find the difference between the number of correct answers in the top group (**RT**) and the number of correct answers in the bottom group (**RB**). Divide this by the total number of people in the top group (**NT**):

$$\frac{RT - RB}{NT}$$

Distribution statistics

Imagine eight students got the following scores:

12 28 19 15 15 35 14 15

The mean

The mean is the sum of all the scores, divided by the number of students:

$$M = \frac{\sum N}{N}$$

M = mean X = the score

\sum = 'the sum of' N = the number of students

$$\frac{\sum N}{N} = \frac{12+28+19+15+15+35+14+15}{8} = \frac{153}{8} = 19.125$$

The mean is therefore 19.13.

The mode, the median

The mode

This is the most frequently occurring score. In this case there are three 15s, and only one of each of the other scores. The mode is therefore 15.

The median

The median is the middle score or the mid-point in the scores. First put the scores in order:

35 28 19 15 15 15 14 12

Then count along to the mid-point of the scores. There are eight students, so the mid-point comes between the fourth and the fifth scores (there are four scores above this point, and four scores below it). Since

the fourth and the fifth scores are both 15, the mid-point is 15. The median is therefore 15.

If the scores had been 35-28-19-17-15-15-14-12, the mid-point would have been between 17 and 15, so the median would have been 16.

The range

The range is the difference between the top score and the bottom score.

The range is: $35 - 12 = 23$

The standard deviation

This is the square root of the average squared deviation from the mean of the students' scores.

$$SD = \sqrt{\frac{\sum (X-M)^2}{N-1}}$$

SD = standard deviation \sum = 'the sum of'

X = the score M = the mean of the scores

N = number of students

Step 1 List scores under X.

Step 2 List the differences between each score and the mean (19.125) under (X - M).

Notes:

(i) Use the exact mean; do not correct to one or two decimal places.

(ii) The numbers under (X - M) should add up to 0.

Step 3 Square all the numbers listed under (X - M) and list under (X - M)²

Step 4 Add up the (X - M) column. The resulting total is

$$\sum (X-M). \quad SD = \sqrt{\frac{\sum (X-M)^2}{N-1}} = \sqrt{\frac{458.91}{8-1}} = \sqrt{\frac{458.91}{7}} = \sqrt{65.56}$$

SD = 8.1

The standard deviation

Intervals of reliability coefficient	Test characteristics
0,90 – 0,99	Excellent
0,85 – 0,89	Very good
0,80 – 0,84	Good
0,75 – 0,79	Satisfactory
0,70 – 0,74	Hardly satisfactory

Kuder Richardson 21

The formula is:

$$\bar{r}_{tt} = \frac{nv - M(n - M)}{(n - 1)v}$$

r_{tt} = the reliability index

n = number of items

v = test variance

M = mean score

This example uses the test results given in *Split half reliability* part above.

1. The number of items in the test is 20, so: n = 20.

2. The mean score is , $\frac{97}{10}$ so: M = 9.7

3. v = 26.9 (To calculate the variance follow steps 1-6 in the *Standard deviation* example above.)

4. nv = 20 x 26.9 = 538

5. n-M = 20 - 9.7 = 10.3

6. M(n-M) = 9.7 x 10.3 = 99.91

7. n - 1 = 20 - 1 = 19

8. (n - 1)v = 19 x 26.9 = 511.1

Replace the symbols in the formula with the numbers:

$$r_{tt} = \frac{538 - 99.9}{511.1} = \frac{438.09}{511.1} = .857$$

The KR21 reliability index is .86.

ДОДАТКИ

Структура навчальної дисципліни

Назви змістових модулів і тем	Кількість годин											
	денна форма						заочна форма					
	усьог о	у тому числі					усьог о	у тому числі				
		л	п	ла б	ін д	с.р .		л	п	ла б	ін д	с.р .
1	2	3	4	5	6	7	8	9	10	11	12	13
Модуль 1												
Змістовий модуль 1. Теоретичні основи мовного тестування та оцінювання												
Тема 1. Предмет і завдання курсу. ЗЕР.		2	1			2						
Тема 2. Основна термінологія з тестування.		2				2						
Тема 3. Основи мовного тестування й оцінювання.		2	1			2						
Тема 4. Специфікація тесту. Конструювання тесту.		2	1			2						
Тема 5. Тестування читання. Тестування аудіювання.		2	2			6						
Тема 6. Тестування		2	1			2						

загального використання мови													
Тема 7. Європейський досвід написання тестових завдань. Коментар до тесту.		2	1			2							
Тема 8. Статистичний аналіз тесту.		2	1			2							
Індивідуальні завдання						10							
Разом за змістовим модулем 1		16	8			30							
Модуль 2. Залік «Практичні навички та вміння складання і презентації тестових завдань, оцінювання якості їх змісту»													
Усього годин	54	16	8	-		30							

Самостійна робота

№ з/п	Назва теми	Кількість годин
1	Предмет і завдання курсу. Основна термінологія з тестування. Основи мовного тестування й оцінювання.	6
2	Специфікація тесту. Конструювання тесту.	2
3	Тестування читання. Тестування аудіювання. Відбір текстів для читання (аудіювання)	6

	відповідно до рівнів Загальноєвропейських мовних рекомендацій».	
4	Тестування загального використання мови.	2
5	Європейський досвід написання тестових завдань. Коментар до тесту.	2
6	Статистичний аналіз тесту.	2
7	Підготовка до заліку. Виконання індивідуальних завдань	10
	Разом	30

Індивідуальні завдання

1. Складання тестового завдання формату «множинний вибір» з читання, аудіювання, загального використання мови (за вибором студента) відповідно до європейських стандартів (10 балів), його оформлення (5 балів).
2. Критичний аналіз тестового завдання колеги за схемою (див. Методичні рекомендації, п.9) (10 балів).

Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота								Сума	
Змістовий модуль №1							Модуль 2		
T1	T2	T3	T4	T5	T6	T7	T8	Залік	100
5	5	10	10	15	10	10	10	25	

Тема 1. Усне опитування – 5 балів.

Тема 2. Тестове завдання – 5 балів.

Тема 3. Усне опитування – 5 балів; Письмове завдання – 5 балів.

Тема 4. Усне опитування – 5 балів; Аналітичне завдання – 5 балів.

Тема 5. Усне опитування – 5 балів; Тестове завдання – 5 балів; Письмове завдання – 5 балів.

Тема 6. Усне опитування – 5 балів; Письмове завдання – 5 балів.

Тема 7. Усне опитування – 5 балів; Аналітичне завдання – 5 балів.

Тема 8. Усне опитування – 5 балів; Аналітичне завдання – 5 балів.

Критерії оцінювання

Система контролю знань студентів здійснюється через:

- **поточний контроль** – (1) усне опитування на лекціях і семінарських заняттях; (2) тестові завдання формату «встановлення відповідності», (3) письмові завдання, (4) аналітичні завдання, які мають на меті перевірити рівень теоретичних знань з теми семінарського заняття;
- **підсумковий контроль** – залік: два індивідуальні завдання (див. п. 9), які демонструють практичні вміння застосування студентами теоретичного матеріалу змістового модуля 1.

Усне опитування реалізує поточний облік успішності студентів на основі чотирибальної шкали оцінок за виконання того чи іншого виду завдання. Так,

- **5 балів** студент отримує за вичерпну відповідь з теоретичного питання, яке розглядається, за умови її оформлення відповідно до норм усного англійського мовлення.
- **4 бали** – за повну відповідь, проте з однією-двома змістовими неточностями, незначною кількістю мовних помилок.
- **3 бали** – за неповну відповідь, або таку, що містить ряд змістових неточностей і незначну кількість мовних помилок.
- **2 бали** – за часткову, з грубими змістовими недоліками відповідь, або за її невідповідність питанню чи відсутність взагалі; відповідь, що містить значну кількість помилок, які порушують акт комунікації, унеможливають розуміння змісту висловленого.

Тестові завдання: 10 завдань формату «встановлення відповідності», виконання яких передбачає вибір правильного варіанту відповіді на питання теми, що розглядається на семінарському занятті. Правильний варіант відповіді оцінюється в

0,5 бала. Нерозбірливо написана, невірна відповідь, її відсутність – 0 балів.

Письмові завдання: 10 запитань з теми, які передбачають надання студентом письмової відповіді, правильність яких перевіряється за ключем. Кожна вірна відповідь оцінюється в 0,5 бала. Невірна відповідь, її відсутність – 0 балів.

Аналітичні завдання: 10 завдань, які передбачають надання студентом відповіді як результат аналізу та узагальнення теоретичних знань з теми, правильність яких перевіряється за ключем. Кожна вірна відповідь оцінюється в 0,5 бала. Невірна відповідь, її відсутність – 0 балів.

Шкала оцінювання: національна та ECTS

Сума балів за всі види навчальної діяльності	Оцінка ECTS	Оцінка за національною шкалою	
		для екзамену, курсового проекту (роботи), практики	для заліку
90 – 100	A	відмінно	зараховано
80 – 89	B	добре	
70 – 79	C		
60 – 69	D	задовільно	
50 – 59	E		
26 – 49	FX	незадовільно з можливістю повторного складання	не зараховано з можливістю повторного складання
0-25	F	незадовільно з обов'язковим повторним вивченням дисципліни	не зараховано з обов'язковим повторним вивченням дисципліни

ЗМІСТ

ВСТУП	3
ЧАСТИНА 1	
ТЕМАТИКА ЛЕКЦІЙ.....	
ТЕМАТИКА ПРАКТИЧНИХ ЗАНЯТЬ.....	
Рекомендована література	
Основна.....	
Допоміжна.....	
Інформаційні ресурси.....	
ЧАСТИНА 2	
ДОВІДКОВІ МАТЕРІАЛИ.....	
Glossary.....	
General Guidelines for Item Writing.....	
Guidelines for Reading Paper.....	
Guidelines for Use of English Paper.....	
Guidelines for Listening Paper.....	
Checklist.....	
Requirements for multiple choice item writing.....	
Guidelines for Writing Paper.....	
Guidelines for Speaking Test.....	
Guidelines for Writing Paired Tasks.....	
General Guidelines for Designing Role-play Tasks.....	
CEFR Descriptors.....	
Feedback Grid.....	
Methods of Test Analysis.....	

ДОДАТКИ

Структура навчальної дисципліни.....

Самостійна робота.....

Індивідуальні завдання.....

Розподіл балів, які отримують студенти.....

Шкала оцінювання: національна та ECTS.....