مجلة أكاديمية شمال

أوروبا المحكمة للدراسات

والبحوث التربوية والإنسانية

ـ الدنمارك

العدد : 22

13/01/2024

تقييم أختبارات اللغة الانجليزية لطلاب مسار العلوم بكلية علوم التمريض. دراسة حالة طلاب الفصل الدراسي الثالث، الدفعة التاسعة عشر للعام الدراسي (2021- 2022م) بجامعة الرباط الوطني في السودان

Assessing the English Examination for Science Track Students. A case Study of Faculty of Nursing Science, Batch (19), Semester (3), During the Academic year (2021 – 2022) at The National Ribat University, Sudan

Prepared by

Asso.Prof.Dr Abdulghani Eissa Tour Mohammed
Applied linguistics, Department of English and Translation, College of Science and Arts, Ar Rass, Qassim University, Saudi Arabia
mohammed@qu.edu.sa

Dr. Ahmed Abdalla Saeed
College of Science and Arts, Baha University, Almikhwa, Saudi Arabia
Ahmad_saeed604@yahoo.com

Dr.Eatdal Balla Eltom Ali
The National Ribat University & Academic Engineer Science and Medicine, Sudan
eatedalblla@gmail.com

Abstract

This study set out to determine if the English test for the third semester or batch (19) of the National Ribat University, Faculty of Nursing Science in Sudan, was given in a way that adequately represented the subject matter and included a sufficient sample of the goals listed in the syllabus during the academic year (2021 – 2022). Descriptive research was used to accomplish this goal, while purposive sampling was carried out regarding the selection of the English exam papers. The data for the analysis was gathered from exam papers and syllabuses, while the English Examination question sheet served as the study's subject. In addition, the test items were analysed using the English syllabus. Afterwards, the data was retyped, and the syllabus was compared. Subsequently, each test item was examined individually to determine the percentage. The researchers finally noticed that not all of the syllabus was covered by the exam questions when invigilating the English language section. Furthermore, it was also clear that some textbook topics were underrepresented in the English language exam. Thus, the lack of representation prompted the researchers to look into the exam's content validity.

Keywords

1. Introduction

Exams for English language proficiency in Sudanese universities have raised concerns. The goals of English language instruction have recently been modified to serve domestic requirements better and keep up with contemporary developments.

Exams are quality control instruments that we use to gauge how well the national educational goals are being met. Practical assessments have the power to assist instructors in improving their methods of instruction and students in modifying their study plans. Exams are another benchmark that can help decision-makers assess how well their kids are doing. Testing is the primary and most significant component of the educational process, according to (Wall, 2005).

The traditional definition of validity in testing and assessment is determining whether a test measures what it is supposed to measure accurately (Hughes, 2003: 26) or determining whether a particular test or any of its parts is appropriate as a measure of what it is supposed to measure (Henning, 1987: 170). The extent to which a test or assessment tool assesses every facet of the subject, idea, or behaviour it is intended to measure is known as content validity. It is described as every effort to demonstrate that the test's material should represent the content under study. To determine whether a test question targets a trait that the instrument is intended to cover, content validity measurement entails analysing each question individually and consulting with specialists. In this process, the test's objectives and the theoretical characteristics of the construct are compared. According to Carroll (1980: 67), defining the learners, assessing their needs, and tailoring the content to meet those needs are the steps necessary to achieve content validity in the assessment of English for Academic Purposes (EAP). Furthermore, Fulcher states that the core idea in establishing test validity in early approaches in the communicative language testing, depends on how best to sample from needs and the target domain" (p. 222–223).

1.1 Statement of the problem

The observed limitation of some question types and the excess of other kinds was the primary motivation for researching the content validity of the English Examination in the third

semester of the National Ribat University's Faculty of Nursing Science. When invigilating the English language segment of the exam, the first co-author of the research team observed that not all of the syllabus was covered by the exam questions. It was also evident that several textbook themes were not adequately covered in the English language exam. The researchers therefore investigated the exam's content validity in response to the dearth of representation.

1.2 Significance of the Study

Teachers' awareness of language testing may increase as a result of the study's findings. Additionally, the study motivates students to engage with the textbook's material. It also evaluates whether or not the test material appropriately reflects the subject areas that need to be examined by going over how the textbook's content categories are divided up among the practical exercises. Lastly, it is helpful for researchers to have a foundational understanding of what makes a good test and for all teachers to know in order to undertake additional research.

1.3 Objectives of the Study

This study aims to achieve the following objectives:
1. To assess the degree of alignment between all language topics covered in the syllabus and the English tests.
2. To assess the relationship between course objectives and the characteristics of English exams.
3. Investigating English exam characteristics from a content validity standpoint.
4. To assess the degree of correlation between the elements of the curriculum and those in the English exams.

1.4 Questions of the Study

1. How do the English exams at Ribat University's Faculty of Nursing Science, third semester, demonstrate the content validity?
2. How do the goals of the course relate to the characteristics of English exams?
3. How closely do all of the language areas listed in the syllabus align with the English examinations?
4. How do the items on the English examination relate to the syllabus?

1.5 Limitation of study

The study is restricted to batch 19, semester three of the National Ribat University's Faculty of Nursing Science, and conducted during the academic year (2012 -2022). The participants come from a single university, hence the outcome cannot be broadly applied.

2. Literature review and previous study

Exams for English language proficiency in Sudanese universities have raised concerns. The goals of English language instruction have recently been modified to serve domestic requirements better and keep up with contemporary developments.

2 .2-Testing

Testing is one of the assessment tools which we can use in the process of teaching and learning. The test is therefore, created based directly on an analysis of the student's English language demands. According to Huge (2003), the test is a tool designed to analyze students'

English language needs and quantify their language proficiency. As stated by Reynolds (2005), a test is defined as "an instrument or steps in which a sample of an individual action is obtained and scored using the basic standard". Although the term "test" has many different meanings, Gronlund and Linn (1990) defined it as a tool or routine process used to measure a sample of behaviour. Based on the previously described definitions, the author recognised that an exam is a specific assessment tool, which we can use for highlighting students' learning and motivation, by assigning one or more assignments. Feedback to the students may also come from the evaluation. The learning process uses assessments for a variety of goals.

2-3 Purpose of Test

Tests are used for a variety of purposes, including determining how well students have met course objectives and gauging their language proficiency. Huge (2003) clarified, "It is difficult to see how rational educational decisions can be made without tests, which will also be needed to provide information about the achievement of groups of learners." Another exciting feature of tests is that, according to Huge (2003), "Tests assist educators and employers in considering their future goals and identifying activities that will help them get there." Numerous kinds of tests are used in the learning process, including achievement, placement, diagnostic, and proficiency exams.

2-4 Types of Tests

Achievement assessments have a direct bearing on language courses since they measure how well students—individually, in groups, or throughout the course—have met learning objectives. There are two types of these tests: progress achievement tests and final achievement tests. Learning outcomes and the instructional materials should be included in achievement exams. Since there are hundreds of facts and terminologies, we expect students to learn in the time allotted, but practically it is impossible to have all the questions we would like to test. However, we can include comprehensive questions in the test that can allow to measure different aspects of the curriculum. Big (2003).

Tests of final achievement are also given at the conclusion of a course of study. Members of teaching institutions, official examining boards, and ministries of education may write and administer them. There is debate among language testers on the extent to which the content of these tests should be connected to the courses they are meant to assess.

Some testers believe that the content of a final achievement test ought to be directly drawn from a thorough course syllabus or from the books and other resources utilized. This strategy is known as the syllabus content approach. According to Huge (2003), the exam's content should be based on the material covered in class since "it only includes material that is believed to represent what the students have actually encountered, making it, at least in this regard, a fair test."

2-5 Construction of Test

Indeed, to construct an accurate test, we need to consider a trained designer. As stated by Farikah (2016) "Creating a set of tests is not as simple as it seems; there must be many qualifications or requirements to consider to make a good test with good quality for the students," .Teachers need to use an excellent exam to ensure accurate measurement. They have a difficult task ahead of them because a successful test must have several qualities, such as validity, reliability, objectivity, and practicality. Moreover, Brown (1995) stated that "the test is built based on the objective of syllabus" while discussing the test construction process.

For that, the instructor needs to be familiar with the goals and substance of the syllabus. The first step in the assessment process is determining the course material. Every subject covered in the course should be covered in the exam questions, with each topic given the proper weight. Bloom et al. (1956) stated that educational objectives is a fundamental benchmark for evaluation. The educational taxonomy theory used six main classes to address the cognitive objective: 1. Knowledge 2. Comprehension 3. Utilisation 4. Examination 5. Synthesis 6. Assessment.

In addition, creating tests is a real issue for lots of educators. It is a genuine difficulty that they encounter on a regular basis. Teachers' training is necessary to gain better skills in the most appropriate ways to avoid hurting their students—most of the time accidentally, though. But how should one go about doing that? When creating a new test, we need to consider the following: What type of test is it going to be? Claims Hughes (1995:49). Success, ultimate or advancement, competence, assessment or identification? What are the specific goals it seeks to achieve? Which skills are going to be evaluated? How detailed do the results have to be? How precise are the needed outcomes? What is the backwash's significance?

2-6. Characteristics of Good Test

Harris (1990) discussed the qualities of a practical test. The following attributes are essential for a good test: validity, reliability, comprehensiveness, usefulness, and scalability. It needs to have a valid exam, the type of tests that provide assessment tools that adequately examine how well students are achieving the intended learning goals. These goals are explicitly expressed in terms of observable behaviour.

Furthermore, thoroughness is yet another trait of an excellent exam. A good test needs to be thorough enough to cover all aspects of the course material that is included in a single syllabus or almost all of it. Siddiek (2010) stated that, the comprehensive exam must assess the learning outcomes of the curriculum or the target course that the students are being taught. Another characteristic of a good exam is its practicality and scalability, which includes being accessible to score and report as well as practical in terms of the allotted time, exerted effort, financial economy. Brown (2001).

2-7 Syllabus

Based on the students' level and the institution's programme, a syllabus is created. In addition, the syllabus that a teacher makes ought to match the abilities of the students. As stated by Richards (2001) "A syllabus is a specification of the content of a course of instruction and list of what will be taught and tested". However, according to Brown (1995), a syllabus primarily addresses the decisions required to arrange the linguistic material of a course or programme. Teachers can select from a wide range of classes and put the necessary materials while instructing students in a classroom. According to the description given, a syllabus is a teaching content that contains that includes a wide range of topics through which students will be taught, and then evaluated on in order to gauge their comprehension of the subject they have studied based on the syllabus. Indicator is crucial to the syllabus's content since it describes the level of competence that students must achieve.

2-8. Syllabus 'objectives

An alternative strategy is to base the test content on the curriculum objectives directly. There are many benefits to this. It first requires course designers to state their goals clearly. Furthermore, it enables exam results to demonstrate the extent to which pupils have met those

goals. This situation consequently increases the pressure on individuals in charge of creating the syllabus and choosing the books and other materials to make sure they are in line with the goals of the course.

Objective-based tests disrupt the cycle of subpar instruction, something that course-content-based tests, almost like a plot within itself, do not. It is much better to base test content on course objectives because this will yield more accurate data regarding student performance on both an individual and group level and will likely have a positive teaching ripple impact.

Focused on the goals of the course, according to Cliff (1981:27). The primary factor contributing to teaching inefficiency and effort dissatisfaction in the classroom is aimlessness. It is especially true when it comes to assessment; how can we, as educators, create an exam or set an assignment if we still need to determine the skills we want the students to learn? As a result, for the assessment to be appropriate, it needs to achieve the goals that have been established for the course. These goals could only be to impart a body of information or abilities, like problem-solving techniques, or they could be to help pupils develop particular beliefs or modify specific behaviours. Thus, the teacher's primary responsibility is to study the course goals before creating the test or task and then craft questions that are based on the content covered in the syllabus.

2-9. Material and Syllabus

The teaching and learning process revolves around the materials, which are essential components that educators must master before imparting them to their pupils. The teacher's materials must match the content of syllabus in order to achieve appropriate learning outcomes. As stated by Brown (1995), materials are any systematic explanation of the method and activities that are to be used in classroom instruction. To ensure that the materials presented in the classroom are adequately retained, a teacher's explanation of the teaching and learning techniques they use and the numerous exercises they assign as test questions must be organised methodically.)

Similar to this, the created test must be guided by the curriculum and measure the ability that the teacher is supposed to be teaching—in this example, language proficiency. Students find it more challenging to respond to tests that lack content validity.

2-10. Validity'

According to Hughes (1989: 22) and Henning (1987:170), "appropriateness of a given test or any of its parts as a measure of what it is purposed to measure" are two definitions of validity in testing and assessment. The former refers to determining whether a test "measures accurately what it is intended to measure." According to this perspective, validity inquiry entails determining if a test "actually does measure" the intended subject. It also assumes that when we construct a test, we have the desire to measure something and that the "something" is "real." We will be challenging these presumptions, which were ingrained in the terminology of validity studies from the first.

Furthermore, validity Validity can be computed technically using Sax (1980:258), which is actual variance divided by achieved variance. Since genuine variance cannot be calculated directly, it must, of course, be estimated in practice. We can define validity in a few different ways. "When a test measures what it says it will measure, it is valid." Garrett (1964:30). "When a test "measures what it ought to measure," it is considered legitimate. Ebel: (1972:436.) "Is the test measuring the things it's supposed to measure? It is a fair test if it does. Lado: (1975:30) "A test needs to be legitimate in order to be considered reasonable. The

accuracy of the test results collection, that it measures   purports is the fundamental definition of validity. Every description listed above places pressure on teachers to assess the students they instruct in courses.

2-11. Types of validity

There are various types of validity:

a. Construct validity
b. Content validity
c. Face validity
d. Criterion measure
e. Criterion-related validity
f. Empirical validity
g. Predictive validity.

Two of these kinds of validity are specifically pertinent to language research. The degree to which a test, questionnaire, checklist, rating scale, or other data collection tool effectively and sufficiently evaluates the specific skills or behaviour it sets out to examine is known as content validity. When it comes to Criterion-Referenced Tests, content validity is especially crucial because the test's content needs to match the material covered in the course as clearly states by Ibrahim Mohamed Alfaki (2012)

Moreover, content validity A test is considered to have content validity, according to Hughes (1995:27), if the items on it comprise and accurately represent a sample of the language abilities, structures, and other topics it is intended to cover. When a test   measures what it intends to measure, it has face validity. When the test provides adequate assessment on achieving pupils' intended learning objectives, which explicitly expressed in an observable behaviour—it is considered legitimate.

Furthermore, the adequacy of the items in the content on the final exam determines content validity. Test success is greatly dependent on the content validity of the sample of language abilities and structures. The student would be under pressure to read every content in his syllabus when multiple course components and skills are covered in the test questions. Additionally, a comprehensive test questions help teachers concentrate on teaching the designated subject for a selected course at the specified time. Exam content allows students and teachers to stay on course by adhering to a syllabus that subject matter experts have carefully crafted to support both individual and community education goals.

Additionally, content validity is an essential tool for the skills assessments that is commonly used in students' evaluation studies. According to Anderson (1975), content validity is the degree to which the test that is employing may effectively measure the traits or dimensions that we want to gauge. According to Chair (2003), content validity is the degree to which test questions accurately reflect the knowledge and abilities in the given subject. A standard method for evaluating content validity is to look at the design and construction process of the test.

2-12. Table of Specifications

According to Walelign (2006:62), the experience of the instructor determines how to create a table of specifications and assign relative weights to the subject's educational objectives. Planning and analysing the material for an achievement test is essential. Any educational material's achievement test must be represented by analysing its content to determine how

accurate the particular test is done by computing and then deciding on a relative weight based on an amount of time spent in teaching each subject.

2-13. Purpose of Table Specifications is

In order to achieve the goals of education, most crucial purpose of the table specifications is to ensure that the test is balanced and that the sample measurement reflects objective of the course and achieve learning outcome of the subject content that it intended to measure in the achievement test. The development of the mind, body, and soul is embodied in educational aims. It also includes the goal of favourably influencing learners' attitudes to achieve academic goals. Siddiek(2010).

2-14 Benefits of the Table of Specifications

With the amount of work put into teaching each subject, the table of specifications aids in creating a fair test simply because it assigns an actual weight to each component of the topic; each topic merits inquiry based on topic's significance. Course specifications also support in the systematic selection of an appropriate teaching content. The method gauges accomplishment's degree to a considerable extent; the instructor assigns questions at various difficulty levels in relation to those objectives. It dramatically boosts the student's trust in the fairness of the test, which will help the applicant manage his time. (Haramaya University, 2017, Ethiopia).

2-15 Component of the table of specifications

There are some elements listed below could be included in the table: a proportionate importance of each subject and vocabulary that will be gauged by how well students succeed academically. The proportional importance of the goals to be assessed. The number of questions for each topic and the level of the cognitive objectives in the test will be determined by the relative weights. Their relative weights will evaluate the test's question(s) for each level of objectives in Ethiopia. Similarly, Walelign (2006) discussed the importance of keeping specific ideas in mind when creating a table of specifications. When creating a table of specifications, we need to take into account the following principles:
- The subject's nature.
- The educational goals.
- Allotted time needed to teach each course based on the characteristics of learners in relation to their degree of study.

2-16 Backwash and Lack of Content Validity

In the study of applied linguistics, backwash is a crucial idea. It alludes to the potential effects and influences that the test may have on the process of instruction and learning. This impact may be advantageous or unfavourable. The validity of the test's content is closely related to the backwash idea. When test's validity is compromised, both teachers and students will suffer as a result. Teachers would have a far better opinion of testing if it consistently had a positive impact on instruction (Siddiek 2010).

According to Weir (1995), students are also able to perform well in language regions or skills where assessments are focused; in other words, students study and practise more of the language areas and abilities that are tested more heavily. By saying this, Weir suggests that learning a language is facilitated when all of the syllabus material is taken into account proportionately during assessment. Finally, test results that demonstrate topic validity—a measure of satisfactory sampling—can be used to estimate students' actual performance

levels. In other words, they promise to produce credible conclusions regarding students' language competency at a specific grade level. Communicative Language Testing, G. J. Weir, 1995. New York: Prentice Hall.

2-17 The effect of lacks content validity of test

In the event that the test is not valid, content standards—which outline precisely what students should understand in a given school year—or   state-created curriculum either authored by teachers or the writers themselves—must be the starting point for achievement test item creation. Developing test items that assess the key competencies and knowledge acquired at a particular grade level is the aim of item writers. The grade-level content standards dictate how many and what kind of exam items are written (Siddiek 2010).

2. 18 Previous study

The Sudan School Certificate English Examination, as it was administered, needs to meet the criteria for being considered a standardised accomplishment test, according to Siddiek (2004). Additionally, it was discovered that the format of the Sudan School Certificate English tests (SSCE) needed to be clarified because it omitted most of the content that pupils were expected to learn from their textbooks. This lack of clarity led to the current scenario where the content validity of the school English exams could be better, which we will look into in this paper. According to Siddiek (2010), the SSCE was a proficiency test rather than an achievement test; therefore, instead of testing the content in the textbook, it concentrated on a broader range of general knowledge of English abilities. The exam's format is comparable to that of TOEFL or IELTS since it addresses language proficiency in its entirety rather than just syllabus topics.

Etika discovered the empirical proof of the English summative exam content validity produced by the expert team for the first-year students' odd semester. The study's conclusions demonstrate the poor content validity of the English summative test items given to first-grade pupils at SMA Dua Mei during their bizarre semester. The content analysis and the percentage both demonstrate it. She concluded that the English summative test is 72% valid in terms of conforming to the indicator based on the percentage. However, the Seminar discovered that just 40% of the test items on SMPAl-Zahra Indonesia during the even semester of 2010–2011 adhered to the English syllabus, indicating poor content validity. Put another way, because a large number of materials were absent from the test, it did not accurately reflect the learning objectives of the semester.

In order to determine whether the coverage of the English language tests given in the college fairly represents the coverage of the textbooks, Asmare (2008) discusses the content validity of the three years' worth of teacher-made achievement tests (TMTs) of English language at Hawassa College of Health Sciences. She does this by observing the contents of the textbook or syllabi and sample test papers. The study's conclusion demonstrates that the sample test papers' content does not accurately reflect the material covered in the textbooks.

According to Siddiek (2010), if a test lacks content validity, it indicates that the items do not cover the topics covered in the syllabus. As a result, the educational goals of the syllabus will not be met, and instruction will be done for no purpose at all. More actual teaching was needed because the majority of the teachers' efforts were focused on getting the kids to pass the tests and get the grades. The money would have been wasted since textbooks were ignored and not read. It's easy to conceive a book that is written but not intended to be read, so why would the Ministry of Education publish and provide thousands of copies to students? Why

did we have to pay for publishers, specialists, and educators when the intended audience—in this case, Sudanese secondary school students—would not utilise or refer to the textbooks?

Only 40% of the exam items at SMP Al-Zahra Indonesia's even semester in 2010–2011 complied with the English syllabus, according to Suminar's analysis of the school's content validity. Put another way, because a large number of materials were absent from the test, it did not accurately reflect the learning objectives of the semester. Evaluation of the EGSEC English Examinations' Content Validity Mekonnen Yibrah Haramaya University, Ethiopia (2017). The study's conclusions showed that there was little correlation and a varied degree of relationship between the material in the sample SATs and the textbooks. The study's findings generally indicate that there were issues with the exam papers' ability to sample each component of the language items fairly. Reading and grammar made up the majority of the exams, whereas the assessment system completely disregarded listening comprehension. The regional state's testing specialists and educators have no affiliation with NEAEA. In order to create an item bank, the NEAEA's test creators do not ask local subject matter experts and testing professionals to submit questions to the organisation. Teachers are unable to provide feedback to the relevant bodies in the quality assurance and examination agency regarding the content validity of the agency's tests because of their limited awareness and personal attitudes, which impact their practice of developing content-valid classroom-level assessments. The details about the work that teachers do at the school level are listed below 1. Clapham, C., Wall, D., and Alderson, J. (1995). Language Assessment: Development and Assessment. Cambridge University Press, Cambridge, UK.

Hanik and Fachrurozy's study is the most recent pertinent one. In the sixth grade of the public elementary school in Uduwanudistrict, they conducted research. The results show that
1. The three test formats are reliable.
2. Each test format's level of difficulty is fair.
3. The three formats' level of discrimination is excellent.
4. The multiple-choice format's distracters are mainly effective.

However, the content validity of the materials being tested only covers some of the fundamental competencies of the school-based curriculum.

In 2013, Ipek Ozera, *, Diana Garveya, Shawn M. Fitzgerald, and Ebed Sulbarana conducted research. The content validity index (CVI) of an English as a Foreign Language (EFL) grade-level test for Turkish primary school students showed low reliability (.52). In contrast, the reliability coefficients for the sub-subsections ranged from 77 to 91, indicating high item reliability.

An Analysis of the Content Validity of English Summative Test Items at the Even Semester of the Second Grade of Junior High School, published in 2014 by the Department of English Education Faculty of Tarbiyah and Teachers Training Syarif Hidayatullah State Islamic University Jakarta, shows that the research findings demonstrate the poor content validity of the English summative test items for the even semester of second-grade students at Al-Amanah Junior High School. The % shows it. The author concluded that the English summative test is 51% valid in terms of conforming to the indicators based on the percentage. Twenty-three items in the test correspond to the indicated indicators listed in the curriculum, and there are 22 items related to conformity. In their 2015–2015 search for the content validity of the summative test items of English for the tenth graders of Sma Negeri 1 Magelang in the school year 2015–2016, Ardhiyan Nugrahanto, Dwi Winarsih, Farikah English Department, Faculty of Education and Teachers Training Tidar University discovered that thirty-two items met the syllabus' requirements, while three items did not. With a rate of 91.42%, this indicates that the English summative test had strong content validity. However,

because one fundamental competency dominated the test items, the representation of the basic skills was not yet proportionate.

3. Methodology

The research approach used for this study is described in this chapter. It explains the population, the instrument, the sample selection process, the methodology, and the data collection processes.

3.1 Study design

The results of the chi-square test of significance and Cramer's V contingency coefficient to the English examination papers have been displayed using quantitative research methods. Additionally, the extent and strength of the association between the syllabus's contents and the contents of the English examination papers have been examined.

3.2 The Research Setting

Since the first co-author works at The National Ribat University, Faculty of Nursing Science, Semester 3 data (syllabus and substance of the English examination papers) were easily accessible, and the study was carried out there. The research was carried out during the academic year (2021–2022).

3.3 The sample

The study's focus was on the Faculty of Nursing Science's syllabus and the four papers from the English Examinations for the third semester. There were (50) questions total across four sections of the English exam papers with (22) as total marks. Section (1) consisted of a comprehension presentation, one passage, and ten true-false objective questions. The grammar-focused sections two, three, and four comprised (40) objective questions (true-false statements, gap-filling items, and multiple-choice questions with (78) as total marks.

3.4.1 Samples and Sampling Technique

The English exam papers from batch (19) were used as a reference. It was selected because of its ease of usage. Furthermore, a convenient sampling technique was used to choose the sample English examination papers because, prior to 2020, teachers had not readily or accessible prepared and delivered the sample exams. The researchers employed two data-gathering instruments to perform the study on the content validity of the tests. These were the curriculum and analysis for the English examination.

3.4. Procedures of Data Collection

The subsequent procedures were conducted in order to evaluate the content validity of the English exams. A list of syllabus objectives, a teacher's handbook, and a textbook containing the items for the English examination were gathered by the researchers, and this is how the task was completed:
1. A list of the objectives for the syllabus and a teaching guide were gathered, and the goals were used to code the materials.
2. Each component of the syllabus's allotted practice task frequency and number of periods were calculated and entered into a table.

3. The anticipated number of questions was ascertained using the tasks allocation table's frequency and compiled periods.

4. Next, the English Examination papers were examined to determine the number of items from each area.

4. Data analysis and desiccation

4.1 introduction

The purpose of this section is to summarise, evaluate, and interpret the study's key findings. the information gathered from the examination of test results and the syllabus's substance.**4.2 Data analysis**

4.2.1syllabus content

Table (1): Summary of total frequency of items in syllabus content areas and sample test papers

Items	Frequency of items in Syllabus content	Frequency of items in test
Passive	7	6
Phrasal verbs	4	-
Active verb	7	4
Future	5	2
Used to do	1	-
Preposition	6	4
For & since	1	1
For, during and while	1	1
Expression with make, do and have	4	-
Reading	5	1
Total	41 = 100%	19 6.3% of the syllabus

This table demonstrates that test papers did not cover slightly less than half of the syllabus's content areas. Reading comprehension, grammar, and vocabulary comprised the three main headings of the course. Each of these main syllabus headings was given a total number of periods, which was then taken and averaged, and their percentages were entered into a table. To determine whether exam papers are entirely or partially representative samples of the syllabus instruction, the number of expected questions from all major headings was computed using these prompts as functional exponents and the number of questions was compared with

the observed questions in the actual exam papers. The following category variables are thought to validate the study and help evaluate the content validity of the English tests.

4.2.2 Course objective

Table (2): Comparison of Course Objectives with syllabus items

Course Objectives	Items of syllabus	Percentage %
Reading	Reading	12.2%
Grammar	Grammar	78.1%
Vocabulary	Vocabulary	9.7%
Writing	Null	0%
Listening	Null	0%
Speaking	Null	0%
Use medical terminology in daily tasks	Null	0%
Total		100%

This table showed that the skills of writing, listening, speaking, and medical terminology of the course objectives should have been covered in the syllabus.

Course Objectives:

By the time this course ends, the learner will have mastered the following:
▪ Read, write, and record in English.
Comprehend written records, conversations, lectures, medical terminology, and library materials.
▪ Write papers, projects, case studies, and sentences.
Use your speaking and listening abilities to communicate in English.
▪ Read and understand English-language materials.
▪ Employ medical jargon in day-to-day activities.

Table (3): Conformity of Test Items with Syllabus

No	Summative Test Items Area	Items Number	Total Frequency
	The conformity test items based on the indicators of English syllabus	19	46.3%
	The inconformity test items based on the indicators of	22	53.7%

	English syllabus		
	Total	41	100

Table 3 shows that the items of the test indicated insufficient items of the test in the syllabus as the relationship between the two variables (here, I prefer to say perfect relationship or strength).

To explain (comparing the percentage with the classifications as follows:

Items of the Test to Syllabus Evaluation	
Range	**Comment**
76% to 100%	Good
56% to 75%	Sufficient
40% to 55%	Insufficient
Less than 40%	Bad

Table (4): Analysis of the syllabus content and test items

Items	Frequency	$X - \overline{X_1} = d$	Mean	$(F d)^2$	Variance (S_1)	t-test	p-value
Passive	7	2.9		412.09			
Future	5	.9		20.25			
Used to	1	-3.1		9.61			
Preposition	6	1.9		129.96			
For &since	1	-3.1		9.61			
During &while	1	-.3.1		9.61			
Active verbs	7	2.9		412.09			

Phrasal verb	4	-.1		.16			
ocabulary	4	-.1		.16			
Reading	5	.9		20.25			
Total	41	0	.1	s=1023.79	25.6		
Analysis of the test items							
Passive	6	3.3		392.04			
Future	2	-.7		1.96			
Preposion	4	1.3		27.04			
For &since	1	-1.7		2.89			
During &while	1	-1.7		2.89			
Active verbs	4	1.3		27.04			
Reading	1	-1.7		2.89			
Total	19	0	.7	s=456.75	76.125	11	.00

Table 4 illustrates that there was insignificant different between syllabus content and the items of the test P-value > 0.05

Table (5): practical exercises in textbooks and test content

Table of specification	Content areas	Observed Frequency	Expected Frequency	O-E	(O-E)2	(O-E)2 /E
Frequencies of practical exercises in textbooks	Language focus (Grammar)	127	123	4	16	.13
	Increase your word power (Vocabulary)	4	3	1	1	,33
	Reading	5	11	-6	36	3.3
	Writing	-	-	-	-	
	Speaking	-	-	-	-	--
	Listening	-	-	-	-	-
Items in test content	Language focus (Grammar)	40	45	5-	25	.6
	Increase your word power (Vocabulary)	-	-	-	-	-
	Reading	10	4	6	36	9
	Writing	-		0		
	Speaking	-	-	-	-	-
	Listening	-	-	-	-	-
Grand total		186		$\sum(O-E) = 0$		$\sum(O-E)2/E= 13.36$
Cramer's Contingency Coefficient (V)						0.07

❖ Large calculated x2 value was evidence that the data were not sampled from the distribution that was expected

❖ Cramer's Contingency Coefficient (V) was close to 0 which indicates a weak association between items of syllabus and items of test.

4.3 Discussion

This study's primary goal was to assess the content validity of English language proficiency tests. The outcomes of the two research instruments—the examination of a textbook and a sample test paper—were examined and discussed. In this section of the study, the researchers have attempted to address the fundamental research concerns, as well as to combine, clarify,

and connect the findings of the data acquired using the two instruments. Furthermore, the study's results were presented in relation to the literature review and received strong support.

The first study question sought to determine whether or not all of the syllabus's content types are covered in classroom exams. The study showed that 46.3% of the exam items were based on the English syllabus, indicating a deficiency in test items compared to the syllabus. The findings of Simachew Gashayea (2019), Suminar (2011), Asmare (2008), and Siddiek (2004), who discovered that the sample exam papers' contents did not fairly reflect the material in the textbooks, corroborate this conclusion. Huge (2003) focused on the need for the test's content to be based on the material covered in the course because it only includes questions deemed to represent what the students have really faced. As a result, it can be seen as fair in this regard.

In response to the second question, the study sought to determine whether or not the course objectives are covered in the course content. 72% of the course goals were outside the syllabus, according to this analysis. Furthermore, Suminar (2010–2011) discovered that certain materials, including speaking and listening, that were left out of the test, meaning that they did not accurately reflect the learning objectives of the seven semesters. In support of the information above, Imrie (1981) stated that all course topics should be included in the test questions, with proper weighting given to each. Arthur Hughes further noted that "there is likely to be harmful backwash if the test content and testing techniques are at variance with the objectives of the course." Concerning the qualities of a practical test, Glenn Fulcher & Fred Davidson (2007) wrote: "A good test must have the following features: - It must be reliable, valid, complete, practical, and scoreable."

In response to question three, the study also sought to determine whether or not the tests had content validity. The researcher discovered that the substance of the syllabus and the exam items differed not significantly (t-test = 0.11 at fd 58 $p > 0.05$), indicating that the material was legitimate. Etika also concluded that the empirical data from the English summative exam for the first year's odd semester was reliable. However, Asmare (2008), Siddiek (2010), and the Department of English Education Faculty of Tarbia (2014) discovered that the school English examinations lacked content validity because the test was not thorough enough in relation to the syllabus content. The goals in the textbooks, which were meant to assist the student at that level, needed to be met as the exam items did not cover the syllabus by testing its components. "A syllabus is a specification of the content of a course of instructions and list of what will be taught and tested," according to a previous writing by Richards (2001). However, according to Brown (1995), a syllabus primarily addresses the decisions required to arrange the linguistic material of a course or programme. Instructors can select from a variety of classes and put the necessary materials in the classroom when they start presenting the content. Another characteristic of an excellent test is its comprehensiveness. A good test needs to be thorough enough to cover all facets of the course material, which is contained in a single syllabus. The comprehensive exam must also meet the learning objectives of the unit or target course that the students are being taught.

Additionally, the table of specifications used chi-square to analyse exam items and practical exercises from textbooks, finding a significant chi-square result of 13. The contents of the test items differed significantly from what was expected of them. A high computed x2 value indicates that the data are not representative of the predicted distribution. This situation leads us to the conclusion that the observed and expected contents of the test items differ significantly. As a result, the material covered in the English exam was needed to represent the contents of the textbooks adequately. For the reasons listed above, a well-designed table of specifications (TOS) is not the foundation upon which the process of creating the test items for the English examination is built.

Additionally, the test's inaccuracy can be attributed to two primary reasons. The first one relates to test methods and content. According to Huge (1999), "If the test is not content-validated, it will have a negative impact on both the teacher and the student (Siddiek, 2010)."

The fourth and last research topic looked at the relationship between the syllabus and test items. The elements in the syllabus and the exam items had a poor correlation, the researcher discovered. The resulting value of 0.07 was in close proximity to zero, signifying a feeble correlation between the syllabus items and the test items administered during the third semester of 2020 at Ribat University's Faculty of Nursing. Furthermore, Simachew Gashayea (2019) discovered in his research that there needed to be a shallow, positive correlation relationship between the test writer and the textbook's content. Cramer's V, which supports the data above, always takes a value in the interval [0, 1]. A high relationship or association between the expected and observed variables is indicated by a value of Cramer's V that is near to one; on the other hand, a weak correlation between these variables is characterised by a value that is close to zero.

5.1 Conclusion

Based on the results of the research of the English examination used at Ribat University, Faculty of Nursing Science-Semester three during the academic year (2021- 2022), the following conclusions are drawn.

1. There needs to be more test items in the syllabus.
2. Most of the course's objectives were not covered by the syllabus.
3. There was a weak association between items of the syllabus and items of the test made V=.07.

5.2 findings

1.   The elements of the assessment indicated insufficient items of the test to the syllabus
2.   Large calculated x2 value was evidence that the data were not sampled from the distribution that was expected
3.   Cramer's Contingency Coefficient (V) was close to 0, which indicates a weak association between items of the syllabus and items of the test.

5.3 Recommendations:

1. The Ribat University Faculty of Nursing Science should implement immediate improvement measures for examinations.
2. Ribat University Faculty of Nursing Science students should pay close attention to striking a balance between the material in their textbooks and their exam materials. To address this, in-service training for the relevant examination bodies is necessary.
3. The number of items should be increased from 110 to 120 in order to reflect the syllabus's aims and contents properly. This idea implies that tests that are longer tend to be more valid and trustworthy than those that are shorter.
4. There is a need for further research in this field.

References:

1.Alderson, J., Claphan, C. And Wall, D. (1995). **Language Test Construction and Evaluation. Cambridge**: Cambridge University Press.

2.Adams, R.J, P.E. Griffin and L. Martin, (1987). **A Latent Trait Method for Measuring a Dimension in Second Language Proficiency**. Language Testing

3.Anderson, Scarvia. B. (1975). **In Encyclopedia of Educational Objectives**. Jossey Bass Publishers. San Francesco. California.

4.Ardhiyan Nugrahanto, Dwi Winarsih, Farikah.(2016).**English Departement**, Faculty of Education and Teachers Training Tidar Univeristy.

5.Bloom et al. (1956) **Taxonomy of Educational Objectives**: The Classification of Educational Goals. Vol. Handbook 1. New York: David Mckay Company.

6. Brown, J. D. (1995). **The elements of language curriculum: A systematic approach to program development. Boston**: Heinle & Heinle

7.Cliff, John C. and Bradford W. Imrie. (1981). **Assessing Students, Appraising Teaching**. Grom Helm ltd. London SW 11David McKay Company. Inc. New York. 176-185.

8.David P. Harris.(2001).*Testing English as a second Language,***New** York: McGraw-Hill Book Company,P. 13. Pedagogy, (New York: Addison Wesley Longman, Inc., p. 86.

9.Ebel, Robert. (1972). **Essential Educational Measurements. Prentice Hall**, Inc. Englewood, Cliff. New York.

10.Ed.Upper Saddle River: Pearson Education, Inc., 2009.

11.Garrett, Henry (1964:30). **Testing for Teachers. American Book Company**. New York. Glossary of Measurement Terms.

12.Gronlund, N.E. (1990). **Measurement and Evaluation in Teaching**. 6th Ed: Macmillan Published Company. New York, London.

13.H.Douglas Brown, **Teaching by Principle an Interactive Approach to Language**

14.Hanik and Fachrurozy. They did a research in 6th grade of elementary school in the public elementary school in Uduwanudistrict.

Hughes, A. (1989). **Testing for Language Teachers. Cambridge**. Cambridge University Pres

15.Hughes, Arthur. (1995). **Testing for Language Teachers. Seventh Printing**. University of Cambridge. Bell &Bain, Ltd. Glasgow.

16.Hughes, Arthur. (2003). **Testing for Language Teachers**. (2nd Ed.). Cambridge University Press, UK.

17.Ibrahim Mohamed Alfaki (2012). **Research Method an Introduction to TESOL** Research Methodology

18.Ipek Ozera, *, Shawn M. Fitzgeralda, Ebed Sulbarana, Diana Garveya (2013): Reliability and content validity of an English as a Foreign Language (EFL) grade-level test for Turkish primary grade students achieved to the Content Validity Index Department of English Education Faculty of Tarbiyah and Teachers Training

19.Jack C. Richards. (2001) **Curriculum Development in Language Teaching**, (Cambridge: Cambridge University Press, p. 2.34

20.James D, Brown (1995). **The Element of Language Curriculum**: A Systematic Approach to Program Development, (New Jersey:Heinle & Heinle Publisher.,1995), p.141.)

21.Lado, Robert. (1975). **Language Testing. Wing Tasi Cheung Printing Company**.

22.Mekonnen Yibrah Haramaya University in Ethiopia (2017). Assessing Content Validity of the EGSEC English Examinations Language Test Construction and Evaluation. Cambridge: Cambridge UniversityPress

23.Reynolds, Cecil R. et al. (2005), **Measurement and Assessment in Education**. 2nd

Sax, Gilbert. (1980). Principles of Educational and Psychological Measurement and Evaluation. Second Edition, Wadsworth Publishing Company. Belmont, California. USA

24.Siddiek, A.G.,(2010). **The Impact of Test content validity on language Teaching and Learning**, Shaqra University, Asian social science journal, 6(12): 133-140.

25.Suminar(2011) found that SMP Al-Zahra Indonesia at even semester 2010/2011

26.Syarif Hidayatullah State Islamic University Jakarta 2014 the title an Analysis On the Content Validity of English summative Test Items at The Even Semester of the Second Grade of Junior High School

27.TKT:Modules 1–3Teaching Knowledge Test, Cambridge: Cambridge University Press .

28.Walelign Admasu, (2006). **Educational Measurement and Evaluation** (Epsy 312). Department of Pedagogical Science, Haramaya University.

29.Weir, G. J. (1995). **Communicative Language Testing**. Prentice Hall, New York