Integrative and discrete-point tasks in EFL tests: a test analysis

Šolaja, Antonio

Repository / Repozitorij:

Sveučilište J. J. Strossmayera u Osijeku

Filozofski fakultet Osijek

Diplomski studij engleskog jezika i književnosti – nastavnički smjer i diplomski studij njemačkog jezika i književnosti – nastavnički smjer

Antonio Šolaja

Integrativni zadaci i zadaci odijeljenih jedinica u ispitima engleskog kao stranog jezika: analiza jednog ispita

Diplomski rad

Mentor: prof. dr. sc. Višnja Pavičić Takač

Osijek, 2016.

Sveučilište J. J. Strossmayera u Osijeku

Filozofski fakultet

Odsjek za engleski jezik i književnost

Diplomski studij engleskog jezika i književnosti – nastavnički smjer i diplomski studij njemačkog jezika i književnosti – nastavnički smjer

Antonio Šolaja

Integrativni zadaci i zadaci odijeljenih jedinica u ispitima engleskog kao stranog jezika:

analiza jednog ispita

Diplomski rad

Znanstveno područje: humanističke znanosti

Znanstveno polje: filologija

Znanstvena grana: anglistika

Mentor: prof. dr. sc. Višnja Pavičić Takač

Osijek, 2016.

J.J. Strossmayer University in Osijek

Faculty of Humanities and Social Sciences

Teaching English as a Foreign Language, MA Programme and

Teaching German as a Foreing Language, MA Programme

Antonio Šolaja

Integrative and discrete point tasks in EFL tests:

analysis of a test

Master's Thesis

Supervisor: Professor Višnja Pavičić Takač

Osijek, 2016

J.J. Strossmayer University in Osijek

Faculty of Humanities and Social Sciences

Teaching English as a Foreign Language, MA Programme and

Teaching German as a Foreing Language, MA Programme

Antonio Šolaja

Integrative and discrete point tasks in EFL tests:

analysis of a test

Master's Thesis

Scientific discipline: Humanities

Scientific field: Philology

Scientific branch: English Studies

Supervisor: Professor Višnja Pavičić Takač

Osijek, 2016

Contents

Summary

The present study examines integrative and discrete-point tasks and compares their efficiency in testing language proficiency. The theoretical part covers language testing and the types of tests that are used in testing language proficiency. It further explains what a standardized language test is and what is being tested. The focus is on the integrative and discrete-point tasks that are part of the entrance exam which is a subtype of a standardized test. The analysis showed that the participants were more successful on the integrative task on than the discrete-point tasks when testing language proficiency.

Key words: discrete-point task, integrative task

Sažetak

U ovom se istraživanju analiziraju integrativni zadatcia i zadatci odijeljenih jedinica te uspoređuje njihovu učinkovitost u testiranju jezične kompetencije. Teorijski dio pokriva jezično testiranje te tipove testova koji se upotrebljavaju u testiranju jezika. Nadalje, objašnjava se standardizirani jezični test te što se njime testira. Težište istraživanja je na integrativnim zadatcima, te zadatcima odjeljenih jedinica koji se pojavljuju u prijamnom ispitu koji je tip standardnog ispita. Istraživanje je pokazalo da su sudionici bili uspješniji u riješavanju integrativnog zadatka nego u riješavanju zadataka odijeljenih jedinica kod testiranja jezične kompetencije vještina.

Ključne riječi: integrativni zadaci, zadaci odijelnih jedinica

1. Introduction

The aim of this research paper is to analyze an English entrance exam from two consecutive years in order to see how different types of tasks test student's language proficiency and whether a single type of task would be enough to test their proficiency level. The main focus is on the comparison of integrative task (cloze test) and discrete-point tasks. The paper sets out with explaining what language testing actually is and continues explaining the importance of language testing when learning a foreign language. In order to develop a language we need constant testing, which in return gives us information on which parts of language need to be worked on and need additional studying.

Language is our main tool for communication. Without it we would be unable to convey our message to other people. It is important to notice that, learning a language and being able to implement it are two different sides of a story. In order to develop our knowledge in a certain language, linguists from all over the world invent new kinds of tests which help us learn and evolve.

This paper consists of theoretical and analytical part. The theoretical part covers the theory of language testing and its importance when learning a language. The analytical part concentrates on analyzing two types of tasks present in English language exams implemented at the Faculty of Humanities and Social Sciences in Osijek.

2. Theoretical part

2.1. Language testing

Teachers spend a lot of time with their students inside a classroom. Due to this most teachers are able to assess their students' progress in any field they are studying. The same goes for language teachers. By spending time with their students, language teachers are able to evaluate the development of their students' language abilities. If this is indeed true, what purpose does language testing play? The purpose of testing, according to Carroll (1961), is to gather information about the knowledge level of the testees which in turn aids the testers in making intelligent decisions about how to further proceed with teaching. In other words, we use tests to determine the learner's current knowledge level and what should be improved. Davidson and Lynch (2002, as cited in Fulcher, 2010: 1) use the term *mandate* to describe the origin of the purpose of the test and suggest that *mandates* can be seen as either internal or external to the institution in which we work. Cronbach (1984, as cited in Fulcher, 2010: 3) stated: "A test is selected for a particular situation and purpose. What tests are pertinent for a psychological examination of a child entering first grade? That depends on what alternative instructional plans the school is prepared to follow. What test of skill in English usage is suitable for surveying a high school class? Those teachers for whom clarity of expression is important will be discontented with a test requiring only that the student choose between grammatically correct and incorrect expressions." According to Douglas (2010) there are many reasons for language tests. Some of the reasons are: fairness, second opinion, instruments of public policy, parental view, etc. Fairness implies that every student has the same chance of success since they all get the same test with the same questions and same conditions. The second reason deals with objectivity. Sometimes, teachers are unable to be completely objective. Tests allow us to get an objective opinion, which in return enables us to confirm our assessment of the student's language level. Instruments of public policy are standardized tests that are used to assess learners across the country, for instance, a college entrance exam. These tests give assessment of the learner's language knowledge level but by same standards as everyone else taking the exam. These tests also enable parents to follow the progress of their children, which is important since parents play a vital role in the development of their children. All language tests consist of specified tasks through which language abilities are elicited. In order to better understand the need for tests, it is important to know the types of tests that are used and what is being assessed when talking about language tests.

2.1.1. Types of language tests

Language tests can be divided into six major categories: language aptitude, language achievement, language proficiency, diagnostic tests, progress test and placement test.

Language aptitude is a potential that a certain person has for learning languages and it is evaluated by using aptitude tests. The language aptitude test, according to Brown (1994) consists of several test items which measure abilities:

- sound-coding ability, i.e. the ability to identify and remember new sounds in a new language
- grammar-coding ability, i.e. the ability to identify the grammatical functions of different parts of sentences
- inductive-learning ability, i.e. the ability to work out meanings without explanation in the new language
- memorization, i.e. the ability to remember and to recall words, patterns, rules in the new language

There are two widely known tools created to measure language aptitude for English: The Modern Language Aptitude Test - MLAT (Carroll and Sapon, 1959) and Pimsleur Language Aptitude Battery - PLAB (Pimsleur, 1966). According to Carroll and Sapon (1959 as cited in Sasaki, 2012) the purpose of the Modern Language Aptitude Test (MLAT) is to predict how well an individual can learn a foreign language in a given amount of time and under given conditions. Carroll believes that given enough time and good instructions it would be possible for anyone to learn a second language but people would differ in terms of speed and ease with which they would learn. The PLAB is a similar test intended for learners from 7th to 12th grade in the USA.

Language achievement tests are used to determine how effective teaching has been or, in other words, how much of what has been taught has been learned. Using an achievement test, we can evaluate the learner's understanding of a specific course or study program. Brown defines an achievement tests as "tests that are limited to particular material covered in a curriculum within a particular time frame" (Brown, 1994: 259).

Language proficiency tests, on the other hand, are used to assess what a person can do with what he already knows. To put it simply, language achievement test deals with specific knowledge or segment of the foreign/second language while the language proficiency test is

used to analyze a person's overall knowledge and his ability to successfully implement it. According to Valette (1977 as cited in Abedi, 2002) the purpose of a proficiency test is to determine whether this language ability correlates with specific language requirements. The Test of English as a Foreign Language, or TOEFL for short, is according to Kunnan (2008) the most well-known and widely used language assessment exam in the world. TOEFL is a standardized test of English proficiency in the USA and is administered by the Educational Testing Service which is located in Princeton. The other well-known proficiency test is the English Language Proficiency Test or ELPT for short. Kunnan (2008) explains that the TOEFL became mandatory for non-American and non-Canadian native speakers of English applicants to undergraduate and graduate programs in U.S. and Canadian English-medium universities.

Diagnostic tests are used to determine in what area the student needs additional teaching and study. An example of a diagnostic test is when at the start of the course, the teacher gives the learners a test to see what areas of language need to be in the syllabus. Prator's *Diagnostic Passage* (Murcia, Brinton and Goodwin, 1996) from 1972 is one of the most known diagnostic tests in the world. It consists of a short written passage that the learner needs to read and is being recorded; the teacher then examines a tape recording of that reading against a very detailed checklist of pronunciation errors.

Most classroom tests are in a form of a progress test. They are used to evaluate the progress that the students make after mastering material taught in the classroom. They also enable students to assess the degree of success of teaching and learning and to see which areas prove more difficult than others.

Placement tests are used to sort new students into teaching groups so that they are approximately at the same level as the others in the group. A good example is one of the more known placement tests: the English Placement Test (EPT) from America. The EPT is designed to assess the level of reading and writing skills of entering undergraduate students so that they can be placed in appropriate courses.

2.1.2. Standardized test – reliability and validity

A standardized test is a test in which all the questions, format, scoring, instructions and reporting of scores are the same for all test takers. It can be divided into two scoring categories: norm-referenced and criterion-referenced tests. In the norm-referenced test, individuals or learners are compared against each other. In the criterion-referenced test, learner's test performance is compared against the performance standard that was pre-determined by the test maker. The purpose of this test, according to Fulcher (2010) is not to discriminate between English speakers but to establish criteria by which a learner can be classified as *operationally proficient*. Stiggins (2008) claims that the purpose of standardized testing is to have large numbers of students write a single test, then to compare any single score against all others to see how an individual's score compares to the large sample. Stiggins continues explaining that the results are then posted on a bell curve that indicates where a score sits within descriptive statistical standards. The most important aspects of a standardized test are:

     a) reliability

     b) validity

Douglas (2010) explains reliability as an accurate measurement of whatever abilities the test is supposed to measure. Lado stated: "Does a test yield the same scores one day and the next if there has been no instruction intervening? That is, does the test yield dependable scores in the sense that they will not fluctuate very much so that we may know that the score obtained by a student is pretty close to the score he would obtain if we gave the test again? If it does, the test is reliable" (Lado, 1961:31). All tests are to a degree inconsistent in their measurement. If a student takes the same test a second time, he or she is likely to get a different score. Douglas (2010) explains that the reasons for this are many: from the instruction being unclear, the test tasks being unfamiliar or too difficult which leads to guessing, the possibility of multiple correct answers, all the way to the test takers being tired or careless.

For a long time, the definition of validity was unchanged. The traditional definition was coined by Ruch (1924, as cited in Fulcher, 2010) who stated:

"By validity is meant the degree to which a test or examination measures what it purports to measure. Validity might also be expressed more simply as the *worthwhileness* of an

examination. For an examination to possess validity it is necessary that the materials actually included be of prime importance, that the questions sample widely among the essentials over which complete mastery can reasonably be expected on the part of the pupils, and that proof can be brought forward that the test elements (questions) can be defended by arguments based on more than mere personal opinion" (Faulcher, 2010: 20). This definition was changed in 1989 by Messick where he stated: "Validity is an integrated evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions based on test scores or other modes of assessment" (Messick, 1993: 3) According to Messick: "the traditional conception of validity divides it into three separate and substitutable types - namely, content, criterion, and constructs validities. This view is fragmented and incomplete, especially in failing to take into account evidence of the value implications of score meaning as a basis for action and of the social consequences of test use" (Messick, 1993: 4). In simpler terms, validity refers to how well a test measures what it is purported to measure. Although a test can be reliable and not valid, in order for it to be a standardized test, it needs to be both. There are many examples of an English language standardized test: Cambridge English Language Assessment, English Language Proficiency Test, Test of English as a Foreign Language, etc.

An example of standardized tests is a college entrance exam. Although different from university to university, the college entrance exam is a proficiency test and is administered by educational institutes such as universities and academies in order to select students for different courses or degrees. Although they can be conducted at any level of education, they are primarily administered at higher levels. Entrance exam across the world are considered a stepping stone into an elite institutions offering professional courses. It is possible for students to pass academic year with minimum marks but in an entrance exam one needs to work really hard in order to break through the competitive environment successfully. As it is with everything else, entrance exams also have their own advantages and disadvantages

Advantages of an entrance exam:

a) Entrance exams act like filters. They enable us to choose capable and hard-working people who would cherish the opportunity to further expand their knowledge

b) Students who are chosen via entrance exam in turn prove helpful in the development of the country.

c) A good performance on the entrance exam becomes an inspiration and encourages students to take part in the competitive environment of these exams.

Disadvantages of an entrance exam:

a) Entrance examinations have enabled the establishment of coaching institutes which are cashing in on the students instead of guiding them properly.

b) Coaching institutions have turned education into business where the motive is to earn as much money as possible. They also teach students about short-cuts instead of encouraging them to gain sufficient knowledge.

c) Entrance exams increase the burden of students. In Croatia, where there is State Matura, the entrance exam is applied only by universities and it varies from university to university.

The exams usually have the same structure but differ in what they test. Language tests for example can test all four language skills or they can focus on two or three.

2.1.3. What is being tested?

The purpose of learning a foreign language is to enable us to communicate with people who use this language on a daily basis. It is important to test whether the learners acquired the necessary skills needed for the communication to be successful. "For all test takers, any test that employs language is, in part, a measure of their language skills. This is of particular concern for test takers whose first language is not the language of the test. Test use with individuals who have not sufficiently acquired the language of the test may introduce construct-irrelevant components to the testing process. In such instances, test results may not reflect accurately the qualities and competencies intended to be measured. … Therefore it is important to consider language background in developing, selecting, and administering tests and in interpreting test performance" (Standards for educational and psychological testing, 1999 as cited in Abedi, 2002).

As seen in the first sentence of the quote above, the purpose of a language test is to test language skills. There are four language skills: listening, speaking, reading and writing. These language skills are divided into two categories: productive and receptive skills. Productive skills are speaking and writing since they require us to *produce* language. Receptive skills, on the other hand, are skills we use to understand the language. Listening and reading fall under this category. We can refer to these four skills as *macro-skills*. In written tests the focus is put on reading and writing skills. An example of a reading skill in a written test would be a cloze test or a reading comprehension task. Writing on the other hand is usually seen in a form of a short essay. Listening skill can also be a part of a written test but requires additional instruments apart from the test itself. It is usually seen as an MCQ task where the testers need to choose an answer based on what they heard on the recording. Apart from *macro-skills* we test other things like for instance grammar, vocabulary or pronunciation. These would be the so called *micro-skills*.

There is a close relationship between all four skills. It is important to develop them equally so that the learner can use the language efficiently. Temple and Gillet (1984, as cited in Aydoğan and Akbarov, 2014), for instance, stated that there is a close relationship between listening and speaking: "Listening cannot be separated from the expressive aspects of oral communication. It is impossible to "teach listening" separately from speaking, or to set aside a portion of the instructional time for listening instruction and ignore it the rest of the time.

Listening is as much a part of group discussions, dramatic play, or puppetry, for example, as the dialogues and actions created. When children develop their communicative powers, they also develop their ability to listen appreciatively and receptively" (Aydoğan and Akbarov, 2014: 673). According to El-Koumy (2002, as cited in Aydoğan and Akbarov, 2014), the focus of teaching is usually a skill-based approach which draws its theoretical roots from behavioral psychology and structural linguistics. This approach is based on five principles:

1. the whole is equal to the sum of its parts

2. there are differences between spoken and written language

3. oral language acquisition precedes the development of literacy

4. language learning is teacher-directed and fact-oriented

5. students' errors are just like 'sins' which should be eliminated at all costs.

The advocates of the skills-based approach view language as a collection of separate skills. Each skill is divided into smaller parts or the so called *sub skills*. The teacher who uses skill-building as his primary focus in learning also uses discrete-point tests (e.g., multiple choice, true or false,) to assess the development of each sub skill before moving to the next (Aydoğan and Akbarov 2014: 674). This being said, most language tests focus on testing discrete-point items.

2.1.4. Integrative and discrete-point tasks/test

Carroll (1961) stated that the purpose of a language test is to determine a person's knowledge and/or ability in the language and to discriminate that person's ability from that of others. In 1961, Carroll introduced the concept of integrative tests and since then there has been an ongoing controversy (Stubbs and Tucker, 1974) between the use of "integrative" and "discrete-point" language tasks/tests (Carroll, 1961; Oller 1972; Rand 1972 as cited in Stubbs and Tucker, 1974) when testing language.

According to Stubbs and Tucker (1974), we can distinguish between two types of language tests: integrative and discrete-point tests. Carroll (1961) recommended tests in which there is less attention paid to specific structure points or lexicon than to the total communicative effect of an utterance. This integrative approach makes less necessary the kind of comparison of language systems upon which much current language test is premised.

Integrative tests, according to Hughes (1989), are tests that require the candidate to combine many language elements in the completion of a task. In other words, they tests language as a whole. They also require of students to combine many linguistic elements and not just focus on one. The concept of communication is such that it cannot be captured in additive tests of grammar, reading, vocabulary, and other discrete points of language. It requires an integrative approach.

Two types of tests have historically been claimed to examples of integrative tests: cloze test and dictation. According to Davies (1978), Oller (1979), and Weir (1988), cloze test can be used to check the readability of textual material for global reading comprehension, investigate the nature of contextual constraints, and estimate overall language proficiency and effectiveness. Taylor (1953) used cloze tests as device for analyzing readability of texts. Oller (1979) has described the procedures of using cloze tests for measuring readability of a text and then the use of these texts to test reading comprehension.

According to Bachman (1990): "cloze procedures do not produce perfect tests of overall language proficiency, they do hold potential for measuring aspects of students' written grammatical competence, consisting of 'knowledge of vocabulary, morphology, syntax and phonology/graphology', and textual competence, knowledge of the cohesive and rhetorical properties of text" (Bachman, as cited in Chapelle and Abraham, 1990: 121).

When scoring a cloze test, there are two types of scoring procedures which are used:

    a) exact word method

    b) acceptable word method

The exact word method demands that the reader guesses the exact word which was used in the original. The acceptable word method, on the other hand, allows the reader to guess and use any word that is acceptable or appropriate in the context.

In addition to integrative test/task, there are also discrete-point tests/tasks. According to Lado (1961), discrete-point tests consist of many questions which try to test a large number of linguistic points, but each question tests only one linguistic point. This theory is constructed on the assumption that "language can be broken down into its component parts and that those parts can be tested successfully" (Lado, 1961). These components include four main skills (listening, speaking, reading, and writing) and various units of language (discrete points) of phonology/graphology, morphology, lexicon, syntax, and discourse. It is believed that a language proficiency test should contain and test all four language skills (listening, reading, speaking and writing) and as many discrete-point tasks as possible. A discrete item approach to teaching language isolates the language and enables teachers and learners to focus on the item itself. For example, it is often useful to practice sounds as discrete items, then in words and connected speech.

Discrete items often appear in testing, where there is a need to focus on knowledge of specific items. Examples of discrete-point tasks are phoneme recognition, yes/no, true/false answers, spelling, word completion, grammar items and multiple-choice tests.

Both integrative and discrete-point tasks have advantages and disadvantages. The advantages of a discrete-point tasks is that they are easy to score and achieve reliable scoring (objective), easily administered & statistically analyzed and can be norm (compared with other test takers) or criterion (reached objective) referenced.

The disadvantages of discrete-point tests/tasks are that they may focus on what test takers know about the language rather than if they can use it, instruction may not go beyond a focus on/ manipulation of language components and they may ignore effects of context.

The advantages of an integrative test/task are that they focus on ability to use language effectively for communicative purposes and using a language involves the integration of all of its features.

The disadvantages are that it is challenging to create clear, meaningful, comprehensive rubrics and level descriptors and they have a potential to be unreliable.

Harmer (2001) stated that the difference between a discrete point-testing and integrative testing is that that discrete-point tasks test on thing at a time (example asking students to choose the correct tense of a verb), and integrative tasks expect that students use a variety of language items – for example, writing an essay.

Stubbs and Tucker (1974) believed that the cloze test, which is a type of an integrative task, represents a great tool for testing language proficiency. Their research showed that cloze test does, indeed, give better overall results than the discrete-point tasks. This research will try to test that hypothesis based on the analysis of an entrance exam.

3. Analysis of a test

3.1. Aim

The aim of this research is to analyze an English entrance exam from two consecutive years and see whether only one type of a task is enough to test a learner's language proficiency. The analysis focuses on two specific types of tasks: the discrete-point tasks and the integrative task. It is hypothesized that participants will be more successful on the cloze test (integrative task) than on the discrete-point tasks as it not only targets different parts of language but it is also more coherent. Although cloze test requires from learners to use linguistic, textual and sometimes world knowledge, it also provides them with contextual advantage.

To achieve that aim, two questions have to be answered:

a) Is there a statistically significant difference in the results on discrete point tasks and the integrative task?

b) Is there a relationship between integrative task and discrete-point task scores?

3.2. Sample

This research is based on an analysis of a college entrance exam administered in two consecutive years: 2008 and 2009. A total of 154 exams (80 from year 2008 and 74 from year 2009) completed by applicants from different parts of Croatia were analyzed. The exam covered the minimum knowledge of English language required to study at the Faculty of Humanities and Social Sciences in Osijek. The test itself consisted of 10 tasks: six gap filling tasks, two MCQ tasks and two transformation tasks. The gap filler tasks were cloze test (integrative task), two tense tasks (the first one dealt with supplying the form of the verb in infinitive, past tense and past participle and the second task involved using a suitable form of the words in the brackets), supply the collocation, prepositions and phrasal verb tasks. The first MCQ task tested reading comprehension and the other MCQ task tested knowledge of English culture. The transformation tasks deal with vocabulary and paraphrasing. Each task had a specific number of points ranging from the lowest of 4 points to the highest of 30 points. The maximum number of points a participant could get on the test was 100. Of all the tasks in the exam, only the first task, which was a cloze test, qualified as an integrative task, while all other tasks were discrete-point tasks. The cloze test required of the participants to use their knowledge of English as it covered areas like verbs (form) as well as phrasal verbs, adjectives, adverbs, collocation, articles but also contextual knowledge. In order to use the right missing word, participants had to understand what they were reading. The context of the text enabled them to narrow the list of possible answers. The discrete-point tasks covered the grammar part but not the comprehension of the text as well.

We can divide all these tasks into four groups:

1) Grammar task: a) Transform (irregular) verbs through tenses (transformation of verbs in infinitive, past tense and past participle)

b) Use the right tense (gap filler and transformation of the given word)

c) Transformation of adjectives into nouns (transformation of the given word)

d) Paraphrase the sentence (transformation)

2) Lexical tasks: a) Use the right collocation (gap filler)

b) Use the right preposition (gap filler)

c) Use the right phrasal verbs (gap filler)

c) Complete the text with the transformation of the given word - vocabulary (gap filler and transformation of the given word)

3) Reading tasks: a) Cloze test (gap filler)

b) Reading comprehension (MCQ)

4) Culture task   (MCQ)

The two exams retained the same structure but there were things that differed. The 2008 exam contained a word formation task which was not present in the 2009 exam. In order to test the same tasks, the word formation task was left out of analysis. Also, since the culture task doesn't test language proficiency but knowledge about English culture, it was also left out of the analysis.

3.3. Procedure

The 154 exams were analyzed task by task. The maximum number of points a participant could get was 100.

Since the tasks varied in the number of points a participant could acquire (lowest 6 and highest 30), the achieved points of the tasks were calculated into percentage so that each and every task would range from 1-100%. This allowed for a comparison of tasks.

For statistical analyses the SPSS statistical programme was used. In addition to descriptive statistics, the following statistical tests were applied: paired sample t-test and the Pearson product-moment correlations test. The paired sample t-test is a statistical technique that is used to compare two population means in the case of two samples that are correlated. By using the paired sample t-test, we can statistically conclude whether one sample shows better results than the other. The Pearson Product Moment Correlation shows the linear relationship between two sets of data.

All discrete-point tasks were taken together and their mean value was calculated. The mean value of the integrative task, which was the cloze test, was compared against the mean value of all other discrete-point tasks. The results of the tests are presented and explained in the section below.

3.4. Results

3.4.1. Difference in results between integrative and discrete-point tasks

Table 1 shows the results of a descriptive analysis of the 2008 exam. The results are presented in percentage. The lowest and highest score received by the participants on each task are shown in the minimum and maximum column. The average score received on the exam is also presented. The .00 score means that the lowest score on the task was 0 points whereas the maximum column shows the best score received on the task.

Table 1. Descriptive analysis of the 2008 exam

	Minimum	Maximum	Mean	Std. Deviation
Cloze test	43.00	97.00	75.88	11.46
Irregular Verbs	.00	100.00	72.37	20.45
Tenses	40.00	100.00	77.12	15.60
Reading MCQ	30.00	100.00	69.75	16.30
Collocations	.00	100.00	58.75	23.56
Prepositions	20.00	100.00	81.25	19.44
Phrasal verbs	.00	100.00	59.00	25.63
Vocabulary	38.00	100.00	84.76	14.72
Paraphrase	.00	71.00	43.58	16.32

Table 2 shows the results of the descriptive analysis of the 2009 exam. The results of the analysis are similar to the 2008 results which in turn will give us better overall results.

Table 2. Descriptive analysis of the 2009 exam

	Minimum	Maximum	Mean	Std. Deviation
Cloze test	43.00	100.00	81.08	11.76
Irregular Verbs	.00	100.00	64.72	22.46
Tenses	40.00	100.00	71.48	14.01
Reading MCQ	30.00	100.00	68.24	15.73
Collocations	25.00	100.00	87.16	17.67
Prepositions	33.00	100.00	70.37	22.18

	Minimum	Maximum	Mean	Std. Deviation
Phrasal verbs	25.00	100.00	74.05	15.66
Vocabulary	.00	100.00	68.24	23.71
Paraphrase	.00	100.00	59.89	20.73

In order to enable easier use of the data found in table 1 and table 2, an additional analysis was made which combined the data from both tables. The results are presented in table 3.

Table 3. Descriptive analysis of the results of 2008 and 2009 exams

	Minimum	Maximum	Mean	Std. Deviation
Cloze test	43.00	100.00	78.38	11.86
Irregular Verbs	.00	100.00	68.70	21.71
Tenses	40.00	100.00	74.41	15.08
Reading MCQ	30.00	100.00	69.02	15.99
Collocations	.00	100.00	72.40	25.27
Prepositions	20.00	100.00	76.38	21.06
Phrasal verbs	.00	100.00	66.23	22.65
Vocabulary	.00	100.00	76.82	21.18
Paraphrase	.00	100.00	51.37	20.25

Table 4 shows the results of the Paired sample t-test of the combined exams. As seen in table 4, the mean value shown is the difference between an average value of the cloze test and all discrete-point tasks. The difference between the integrative task (cloze test) and the discrete-point tasks (irregular verbs, tense, MCQ, collocations, prepositions, phrasal verbs, vocabulary and paraphrase) is statistically significant, $p < .001$.

To answer the first research question which was "Is there a statistically significant difference in the results on discrete point tasks and the integrative task", the answer would be positive.

Table 4. Integrative vs. discrete-point tasks in the 2008 and 2009 exams (Paired samples t-test)

	Mean	SD	t	df	Sig.
Cloze test / Irregular verbs	78.38 68.70	22.25	5.39	153	.000***
Cloze test / Tense	78.38 74.41	17.53	2.80	153	.006
Cloze test / MCQ	78.38 69.02	16.37	7.09	153	.000***
Cloze test / Collocation	78.38 72.40	23.02	3.22	153	.002
Cloze test / Preposition	78.38 76.38	19.50	1.26	153	.207
Cloze test / Phrasal V.	78.38 66.23	20.10	7.49	153	.000***
Cloze test / Vocabulary	78.38 76.82	21.37	.90	153	.367
Cloze test/ Paraphrase	78.38 51.37	17.66	18.97	153	.000***

$***p < .001$

3.4.2. Relationship between integrative and discrete-point tasks

To answer the question: "Is there a relationship between integrative task and discrete-point tasks?" the Pearson product-moment correlations test was conducted. The results of the correlations test (Table 5) show the degree of correlation between the integrative task and discrete-point tasks. Since 8 out of 9 tasks show a positive association between variables, the researcher decided to test whether the variable "tense" from the 2009 exam will show the same results as the one from 2008. Before that however, there is one more thing to take into account and that is whether the correlation between the two variables is significant or not. As shown in table 5, the correlation between most variables is significant. As expected, the "tense" variable isn't significant at all. The exception that appears here is the "vocabulary" variable which also shows that it isn't significant. In order to test if this exception is recurring, the researcher used the same test on the 2009 exam.

Table 5. Pearson product-moment correlations test 2008

	Pearson correlation	Sig.
Irregular verbs	.402	.000***
Tenses	-.041	.715
Reading comp	.337	.002**
Collocations	.350	.001**
Prepositions	.373	.001**
Phrasal verbs	.479	.000***
Vocabulary	.184	.103
Paraphrase	.401	.000***

p < .001

The results of the Pearson product-moment correlations test presented in table 6 show that there is indeed a positive association between the cloze test and all other discrete-point tasks. The significance, on the other hand, seems to be similar to the 2008 exam, but instead of the variables "tense" and "vocabulary", the exception in the 2009 exam is in the variable "irregular verbs". Nonetheless, the results in both tables show that in more than 80% of cases, there is a significant correlation between the variables which is why we are able to ignore the exceptions that appear.

Table 6. Pearson product moment correlations test 2009

	Pearson correlation	Sig.
Irregular verbs	.156	.184
Tenses	.529	.000***
Reading comp	.382	.001**
Collocations	.389	.001**
Prepositions	.595	.000***
Phrasal verbs	.369	.001**
Vocabulary	.541	.000***
Paraphrase	.509	.000***

p < .001

3.4.3. Differences in results and relationship of integrative task and merged discrete-point tasks

The results of the paired sample t-test, after merging all discrete-point tasks into one, show that there is still a significant difference between the cloze test and the discrete-point tasks. This consolidates the results even further. Table 8 clearly shows that there is indeed a significant difference.

Table 8: Paired sample t-test cloze test and discrete-point tasks

	Mean	SD	t	df	Sig.
Cloze test / Discrete -point tasks	78,38 69,42	9.48784	11.723	153	.000***

p < .001

The results of the Pearson product-moment correlations test, after merging the discrete-point tasks into one, are presented in table 6. The results show that there is nonetheless a positive association between the variables.

Table 9: Correlation between cloze test and discrete-point tasks

	N	r	Sig.
Cloze test / Discrete -point tasks	154	.659	.000***

p < .001

The results in tables 8 and 9 strengthen the results gotten in the tables from 1 to 6. The results are further elaborated in the section below.

4. Discussion

The results of the research show that the participants scored higher on the cloze test, which is an integrative task, than on the discrete-point tasks (irregular verbs, tenses, word formation, MCQ, collocations, prepositions, phrasal verbs, vocabulary in context) which appear in the exam. The results were similar in both exams. The reason for this is that although the cloze test requires from learners to use linguistic, textual and sometimes world knowledge, it also provides them with contextual advantage. Since the participants had contextual advantage, it was easier for them to narrow the search for the missing words.

Also, the significance between the differences of the average of the compared variables is very high in all samples. The Pearson product-moment correlations test shows that in most cases there is a positive association between the variables. Furthermore, not only is the average of the cloze test higher than that of the discrete-point tasks, but the degree of correlation between a cloze test and discrete-point tasks is very significant.

According to these results, when testing a student's language proficiency, the hypothesis that the participants will score higher on the integrative task is confirmed. Since communication encompasses using many different parts of language simultaneously, it is this researcher's belief that an integrative test is a perfect tool for testing student's language proficiency. If a student is able to "juggle" different parts of language while speaking or writing in a foreign language, then his language proficiency is obviously high. However, even though the results were in favor of a cloze test, the difference between the mean values ranged in most cases 2-10% which, on a larger scale, isn't as significant when testing a learner's overall knowledge level.

The results may differ, though, if a different type of an integrative task is used, such as an essay. Since an essay requires from participants to write their own sentences, the score might be lower than that of a cloze test which gives us context and requires the usage of a word and not of a whole sentence. It would be interesting to redo the research using an essay as an integrative task and then to compare the results of the essay research with the results of cloze test research.

5. Conclusion

This research analyzed an integrative task (cloze test) and discrete-point tasks against each other in order to see which type of task would better represent a participant's proficiency level. The analysis was conducted on two university entrance exams from two consecutive years and the results show that participants were more successful on an integrative task. The results of the research show that the participants scored higher on the integrative test than on the discrete-point tasks. The difference between the averages of the integrative and discrete-point tasks was very significant. The analysis also confirmed that there is a positive association between the cloze test and discrete-point tasks.

This research succeeded in confirming that in today's world, a cloze test (integrative task) still represents a great tool for testing language proficiency as it did when proposed and used by Stubbs and Tucker (1973), but whether it's enough is hard to say. For a successful communication, a language user needs to be able to use several language units at once which is why an integrative task represents a perfect tool for testing proficiency. That being said, discrete-point task still play a vital role in developing a learner's language knowledge which is why the combination of integrative and discrete-point tasks gives us the best results and enables us to effectively assess learner's language knowledge.

6. Bibliography

Abdul-Jabbar, A. D. (2010). Cloze tests: an integrative approach, University of Al Mustansiriyah. Journal of the College of Basic Education: 105-114.

Abedi J. (2002). Standardized achievement tests and English language learners: psychometrics issues. Educational Assessment, vol. 8 (3): 231–257.

Aitken K. G. (1975). Problems in cloze testing re-examined, TESL Reporter, vol. 8 (2): 9-17.

Alderson, J. C. (1979). The cloze procedure and proficiency in English as a foreign language. TESOL Quarterly, vol. 13: 219-227

Aydoğan H. and Akbarov A. A. (2014). The four basic language skills, whole language and intergraded skill approach in mainstream university classrooms in Turkey. Mediterranean Journal of Social Sciences, vol. 5 (9): 672-678

Brown, H. D. (1994). Teaching by Principles: An Interactive Approach to Language Pedagogy (3rd ed.). Englewood Cliffs, NJ: Prentice Hall

Burke, K. (1999). The Mindful School: How to Assess Authentic Learning (3rd ed.). Arlington Heights, IL: Skylight Publishing

Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency. Testing the English Proficiency of Foreign Students. Washington, D.C.: Center for Applied Linguistics

Celce-Murcia M, Brinton D. M. and Goodwin J.M. (1996). Teaching Pronunciation: Reference for Teachers of English to Speakers of Other Languages. New York, Cambridge University Press

Chappell C. A. and Abraham R. G. (1990). Cloze method: what difference does it make? Language Testing, vol. 7 (2): 121-146

Council of Europe (2001). Common European Framework of Reference for Languages: learning, teaching, assessment. New York, Cambridge University Press

Davidson, F. and Lynch, B. (2002). Testcraft: A Teacher's Guide to Writing and Using Language Test Specifications. New Haven: Yale University Press

Douglas D. (2010). Understanding Language Testing. New York, Routledge.

Farhady, H. (1979). The disjunctive fallacy between discrete-point and integrative tests. TESOL Quarterly, vol. 13 (3): 347-357.

Fulcher, G. (2010). Practical Language Testing. London: Hodder Education.

Hughes, A. (1989). Testing for Language Teachers, Cambridge: Cambridge University Press.

Harmer, J. (2001). The Practice of English language teaching. Essex, England: Longman.

Kunnan, A. J. (2008). Large-scale language assessment. E. Shohamy and N. Hornberger (Eds.) Encyclopedia of Language and Education. (2nd ed.) vol. 7: Language Testing and Assessment: 135-155.

Messick S. (1993). Foundations of validity: Meaning and consequences in psychological assessment. ETS Research Report Series: 1-16.

Oller, J. W. (1973). Integrative and discrete point tests at UCLA: A reply. UCLA Workpapers in TESL, vol. 7: 85 – 94.

Oller, J. W. (1973). Cloze tests of second language proficiency and what they measure. Language Learning: A Journal of Research in Language Studies, vol. 23: 105-118

Perkins, K. (1981), A statistical analysis of different instruments to measure short-term in an L2 immersion program. RELC Journal, vol. 12 (2): 35-49

Pimsleur, P., Reed, D. J. and Stansfield, C. W. (2004). Pimsleur Language Aptitude Battery: Manual, N. Bethesda: Second Language Testing Foundation

Sasaki, M. (2012). The modern language aptitude test (paper-and-pencil version). Language Testing 29 (2): 315-321

Stiggins, R. J. (1999). Are you assessment literate? The High School Journal, vol. 6(5): 20-23

Stiggins, R. J. (2008). An Introduction to Student-involved Assessment for Learning (5th ed.). Columbus, OH: Pearson Merrill Prentice Hall

Stubbs J. B. and Tucker G. R. (1974), The cloze test as a measure of English proficiency. The Modern Language Journal, vol. 58 (5/6): 239-241

Taylor, W. (1953). Cloze procedure: a new tool for measuring readability. Journalism Quarterly, vol. 30 (4): 415-433