Assessing Learners Cognitive Techniques in Language Comprehension Texts

Ms. Ashwaq Sukkar Khayoon.
Research Scholar

Prof. **Dr. Muhammad Ansari**, M.A, B.Ed., L.L.M., M.Phil
Department of Linguistic
Specialization: Phonetics and Phonology.
College of Arts and Social Science
Osmania University, Hyderabad, Telangana, India

ABSTRACT:
Language testing is one of the significant procedures and in language teaching and learning that provides information for teachers and learners and reflects their development in these processes. The teacher utilizes classroom methods and techniques that derive the knowledge of the process of language learning and the interaction between the teacher and the learners, and this can influence testing. Also, aspects of language that are tested to find their ways into the teaching program. This is known as "washback".

Teacher – made tests are on type of the test that reflects this phenomenon hence, the present study aims at :

1. Measuring the sensivity of four techniques of reading comprehension test, namely true – false items, cloze test, multiple – choice and answering question.

2. Identify Iraqi English as foreign language students perception of washback impact.

3. Finding out the relationship between reading comprehension and washback.

4. To achieve the aims of this study three tentative hypotheses are as follows :

*Washback has a positive effect on the process of teaching learning.

*The classical method that is used in teaching reading comprehension do not meet the criteria of good language tests.

*There is statistically significant relationship between washback and reading comprehension.

The random sample of the study consists of 91english as foreign language learners from second year students at Baghdad Secondary School for Girls during the academic year 2016-2017. In order to fulfill the aims of the study and test its hypotheses the researcher has used the procedure of testing, according to the criteria of good language test. The instruments have been exposed to a jury of experts for the purpose of ascertaining their face validity.

INTRODUCTION
PROBLEM AND ITS SIGNIFICANCE

A **test** or **examination** (informally, **exam** or **evaluation**) is an assessment intended to measure a test-taker's knowledge, skill, and aptitude. A test may be administered verbally, on paper, on a computer. Tests vary in style, and requirements. For example, in a closed book test, a test taker is often required to rely upon memory to respond to specific

items whereas in an open book test, a test taker may use one or more supplementary tools such as a reference book or calculator when responding to an item. A test may be administered formally or informally. An example of an informal test would be a reading test administered by a parent to a child. An example of a formal test would be a final examination administered by a teacher in a classroom. Formal testing often1 results in a grade or a test score(Bachman,1996:55). The norm may be established independently, or by statistical analysis of a large number of participants. An exam is meant to test a student's knowledge or willingness to give time to manipulate that subject.

A standardized test is any test that is administered and scored in a consistent manner to ensure legal defensibility. Standardized tests are often used in education, professional certification, psychology and many other fields.A non-standardized test is usually flexible in scope and format, variable in difficulty and significance. Since these tests are usually developed by individual instructors, the format and difficulty of these tests may not be widely adopted or used by other instructors or institutions. A non-standardized test may be used to determine the proficiency level of students, to motivate students to study, and to provide feedback to students. In some instances, a teacher may develop non-standardized tests that resemble standardized tests in scope, format, and difficulty for the purpose of preparing their students for an upcoming standardized test Finally, the frequency and setting by which a non-standardized tests are administered are highly variable and are usually constrained by the duration of the class

period. A class instructor may for example, administer a test on a weekly basis or just twice a semester. Depending on the policy of the instructor or institution, the duration of each test itself may last for only five minutes to an entire class period.

In contrasts to non-standardized tests, standardized tests are widely used, fixed in terms of scope, difficulty and format, and are usually significant in consequences. Standardized tests are usually held on fixed dates as determined by the test developer, educational institution,which may or may not be administered by the instructor, held within the classroom, or constrained by the classroom period. Although there is little variability between different copies of the same type of standardized test, there is variability between different types of standardized tests.

Any test with important consequences for the individual test taker is referred to as a high-stakes test. A test may be developed and administered by an instructor, or a test provider. In some instances, the developer of the test may not be directly responsible for its administration. For example, Educational Testing Service (ETS), a nonprofit educational testing and assessment organization, develops standardized tests such as the SAT but may not directly be involved in the administration or proctoring of these tests. As with the development and administration of educational tests, the format and level of difficulty of the tests themselves are highly variable and there is no general consensus or invariable standard for test formats and difficulty. Often, the format and difficulty of the test is dependent upon the educational philosophy of the

International Journal of Research

Available at https://edupediapublications.org/journals

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 08
July 2017

instructor, subject matter, class size, policy of the educational institution, and requirements of accreditation or governing bodies. In general, tests developed and administered by individual instructors are non-standardized whereas tests developed by testing organizations are standardized. Examinations or tests play an important role in the field of learning-teaching situation. They enable us to know how much to know whether the objectives of teaching the subject have been achieved or not. So tests or examinations are necessary. As a result, the best formula in successful teaching is :

"Teach, test, then ratest and re-test"

The word "examine" is taken from the Latin word 'examen' which means the pointer of balance. That pointer tells by its movements whether the weights in the two pans are equal or not. In an examination, a candidate is weighted as if in a balance. He is compared with some standards. For every examination, standards are fixed by the examining authorities. Then through examination, the standards are compared with those standards.

Testing and Assessment :

Assessment (Or evolution) can be called the technique of measuring the student's growth or progress in learning. It is commonly believed that evaluation is the same s testing. So it is generally supposed that when students are taking a test or exam, we are evaluation them. However, evaluation involves more than just testing student's outcomes. Testing is only one aspect of evaluation. Even the term "assessment" is different from "evaluation". Assessment is made toward the end of a course to find out what the learner has learnt. Thus, it

resembles "examinations" and "test". It has a limited perspective with a focuses on the means. It is intended to serve the learning process. In this way, evaluation has a wider scope as compared to tests, examinations and assessment.

Test has a powerful influence on language learner who are preparing to take these tests, and on the teachers who try to help them prepare. Swain (1985:43) states that it has frequently been noted that teachers will teach to a tests is, if they know the content of a test and/or the format of a test, they will teach their student accordingly.According to Wall and Alderson (1993:41) it is common to claim the existence of backwash (the impact of a test on teaching) and to declare that tests can be powerful determiners, both positively and negatively.

Backwash is often referred to as the influence tests on teaching and learning. In other words, tests affect teachers and learners and therefore affect teaching and learning activities. These effects might be either beneficial or harmful depending on various Factors that have not get defined (identified) yet. However, whether separate and indentifiable phenomenon of backwash" is well document academic phenomenon common to nearly all institutional learning process.

VALUE

It is hoped that the present study will be valueable in developing English as foreign language students perception of backwash impact enhancing their techniques and methods.

THE ORIGIN OF WASHBACK

Washback (Alderson & Wall,1993), together with other similar related terms such as *backwash* (Biggs, 1995, 1996),*test impact*(Bechman& Palmer,

1996; Baker, 1991), *systemic validity* (Frederickson & Collins, 1989), *consequential lvalidity* (Messick, 1989, 1996), *measurement – driven instruction* (Popham, 1983, 1987), *curriculum alignment* (Shepard, 1993) and possibly other terms, all refer to different facets of the same phenomenon. The concept and study of washback has also been derived from recent developments in language testing and measurement driven reform in the areas of general educational assessment. Washback introduced in language testing courses for teachers as a powerful concept that all test designers need to strongly consider, and of which most classroom teachers are all too aware. Davies (1985) asks whether test should necessarily follow the curriculum and suggests that perhaps tests ought to lead and influence curriculum. Morrow (1986:6) further used the term 'washback validity' to describe the quality of the relationship between testing, teaching, and learning. He claimed that '…in essence an examination of washback validity would take testing researchers into the classroom in order to observe the effect of their tests in action'. In general education, researchers have claimed that high-stakes testing might practices or motivate teachers to manipulate students' test scores or it might cause teachers to teach to the test. The above mentioned beliefs in the effects of tests, however, have seldom been empirically corroborated in the area of language testing. It was not until Alderson and Wall (1993) who proposed their washback hypotheses as the foundation for further research, that researchers undertook washback studies. To date, researchers have paid most of their attention to the washback of tests on four domains of teaching practice:

- Content of teaching
- Teaching methods
- Assessment methods and more broadly
- Overall teaching style, classroom atmosphere and teachers' feelings toward the test.

Interest in this important area for teachers, learners, and other stakeholders will undoubtedly grow as tests-especially high stakes tests are used on issues and solving problems inherent in tests in order to increase their reliability and validity.

The definition and scope of washback

Washback is a reference to tests' influences on educational systems and even on society in general. Washback is a term commonly used in language testing, yet it is rarely found in dictionaries. However, the word *'backwash'* can be found in certain dictionaries and is defined as 'the unwelcome repercussion of some social action' by the *New Webster's Comprehensive Dictionary of the English Language* and 'unpleasant after –effects of an event or situation' by *Collin's Dictionary of English Language*. Washback, commonly used in the field of applied linguistics, refers to 'the impact of a test on teaching'(Alderson and Wall 1993). It refers to the extent to which a test influences language teachers and learners to do things 'they would not necessarily otherwise do because of the test'. Messick (1996:241) points out that 'washback, a concept prominent in applied linguistics, refers to the extent to which the introduction and the use of a test influences language teachers and learners to do things they 'would not

otherwise do that promote or inhibit language learning'. Shohamy notes (1992:513) that 'this phenomenon is the result of the strong authority of external testing and the major impart it has on the lives of tests takers.

Biggs(1995:12) uses the term 'backwash' to refer to the fact that testing drives not only the curriculum but teaching methods and students' approaches to learning. Spolsky (1994:55) commented that 'backwash is better applied only to accidental side-effects of examinations and not to those effects intended when the first purpose of the examination is control of the curriculum: According to Alderson and Wall (1993:115) , the notion that testing influences teaching is referred to as 'backwash' in general educational circles, but it has come tobe known as 'backwash' among British applied linguistics, though they see no reason, semantic or pragmatic, for preferring either term.Pearson (1988:98) points out that 'public examinations influence the attitudes,behaviours, and motivation of teachers, learners and because examination often come at the end of a course, this influence is seen working in a backward direction, hence the term 'washback'. He emphasizes that the direction in which washback actually works must be forwards in time. Alderson and Wall (1993:55) also emphasize the fact that evidence of washback is typically demonstrated in behavioral and attitudinal top-down attempt to elicit positive washback effects on the teaching and learning of English.

Manipulation of Washback

The 'washback' or 'backwash' effect of testing is a well-documented academic phenomenon common to nearly all institutional learning processes. The washback effect has been described as: 'the influence of testing on teaching and learning' (Gates: 1995). There is a natural tendency for both teachers and students to tailor their classroom activities to the demands of the test, especially when the test is very important to the future of the students, and pass rates are used as a measure of teacher success. This influence of test on the classroom (referred to as washback by language testers) is, of course, very important…'(Buck: 1988).

Washback is not restricted to learners and teachers. thewashback effect of testing, will be more with the wider social effects of washback. (Bachman and Palmer :1996) consider washback to be a subset of a test's impact on society, education system and individuals. They believe that test impact operates at two levels:

1. The micro level (i.e. the effect of the test on individual students and teachers).

2. The macro level or the impact the test may have on society and the educational system. Testers consider washback as one dimension of impact, describing effects on the educational context (Hamp – Lyons 1997): others see washback and impact as separate concepts relating respectively to "micro" and "macro" effects within society. Most testers locate both concepts within the theoretical notion of "consequential validity" in which the social consequences of testing are part of a broader, unified concept of test validity (Messick 1989, 1996). Consequential validity has been extensively discussed among language testers in recent years (Kunnan 2000). Most testers now acknowledge that washback and impact are highly complex Phenomena: some take a stronger view derived from critical theory in which language testing

is characterized as the exercise of power by one party over another (Shohamy 2001).

Another consideration to be kept in mind as we briefly consider the washback effect of language testing is that test design and content directly affect the validity of a test, i.e. the degree to which a test is measuring what it claims to measure. A test's validity determines the reliability of its results. Some writers claim that a test's validity (the extent to which a test measures what it is supposed to measure) should be established by the degree to which it has a positive influence on teaching. The problem however lies with the evidence of washback: a test might influence what is taught but not how it is taught, might influence teacher behaviors but not learner behaviors, or might influence both with little or no influence in skills.

Dynamics of Washback

Tests are designed to, and are supposed to evaluate what is taught in a formal educational context. They are also utilized to choose some individuals among others for a specific job or for promotion. Whether it is an avhievement or a proficiency test, washback effect emerges the moment learners and teachers shape their behaviors in parallel with the nature of the test to be offered during learning and teaching process. It also emerges when these examinations begin to influence social and economic life in a society. "Public examinations influence the attitudes, behaviors, and motivation of teachers, learners, and parents, and, because examinations often come at the end of a course" (Davies,1990:98) .

(Messick,1996) defines the nature of the washback as "the extent to which a test influences language teachers and learners to do things they would not

necessarily otherwise do that promote or inhibit language learning".(Wall and Anderson,1996:241) claim that tests can influence the classroom practices both negatively and positively

Only recent studies started to empirically investigate the phenomenon of washback. Assertions about the nature, extent, and direction (positive/negative) of impact in language testing have often been based on assumptions rather than on empirical evidence.(Alderson&Wall ,1993:90) argued the need for empirical investigation and were among the first to develop appropriate research hypotheses. Since then, language testers have developed various instruments for measuring washback and impact, and evaluating degree to which they may be considered positive or negative (Saville and Hawkey, 2004:20). Negative washback is said to occur when a test's content of format is based on a narrow definition of language ability, and so constrains the teaching/ learning context. Davies offer the following illustration: If, for example, the skill of writing is tested only by multiple choice items then there is great pressure to practice the skill of writing itself. Positive washback is said to result when a testing procedure encourages "good" teaching practice: for example, an oral proficiency test is introduced in the expectation that it will promote the teaching of speaking skills.

Positive Washback

According to(Messick,1996:242) "For optimal positive washback there should be little, if any difference between activities involved in learning the language and activities involved in preparing for the test"However, creating a positive washback is not an easy task that has certain procedures and steps which can be applied where and when necessary. The complexity of washback

effect lies in the unique nature and dynamics of each educational system and society in which hundreds of different variables play a part in shaping the expectancies of the learners and in their learning choices. "A poor test may be associated with positive effects and a good test with negative effects because of other things that are done or not done…" (Messick, 1996, p.242). Therefore, the attempts of designing test to create a positive washback effect on both learners and teachers are challenged by so many factors operating in social, economic, cultural, and institutional strata of a specific educational system.

Negative Washback

Language tests are often criticized for their negative influence on teaching so called 'negative washback'. (Vernon,1956:166) commented that teachers tend to ignore subjects and activities that do not directly contribute to passing the exam, and claimed that examinations 'distort the curriculum'.(Davies,125:1968) for example ,indicates that 'all the washback effect has been bad; designed as testing devices, examinations have become teaching devices, work is directed to what are in effort-if not in fact-past examination papers and consequently becomes narrow and uninspired. (Alderson &wall ,1993 :5) refer to 'negative washback) as the negative or undesirable effect on teaching and learning of a particular test. In this case, 'poor' usually means something that the teacher or learner does not wish to teach or learn.

The tests may well fail to reflect the learning principles and/or the course objectives to which they are supposedly related.(Fish,1988:30) discovered that teachers reacted negatively to pressure created by public displays of classroom scores, and also found that relatively inexperienced teachers felt greater anxiety and accountability pressure than did experienced teachers.(Noble&Smith,1994:99) pointed out that high-stakes testing affected teachers directly and negatively, and that 'teaching test-taking skills and drilling on multiple choice worksheets is likely to boost the scores but unlikely to promote general understanding'.

Smith concluded from an extensive qualitative study of the role of external testing in elementary schools that 'testing programs substantially reduce the time available for instruction, narrow curricular offerings and modes of instruction and potentially reduce the capacities of teachers to teach content and to use methods and materials that are incompatible with standardized testing formats'. (Heyneman,1987:262) concluded that 'testing is a profession, but it is highly susceptible to political interference'.

To a large extent, the quality of tests relies on the ability of a test agency to pursue professional ends autonomously. If the consequences of a particular test for teaching and learning are to be evaluated, the educational context in which the test takes place needs to be investigated. Whether the washback effect is positive or negative will largely depend on how it works and within which educational contexts it is situated. In a purely academic environment, in which most research has been performed to date, language testing takes place in a largely closed circuit, where success or failure is the key to the next step in the academic chain. Proficiency testing is truly 'high stakes' because its outcome directly engages the real world in terms of safety and career.

ACHIEVEMENT

Achievement measures the amount of academic content a student learns in a determined amount of time. Each grade level has learning goals or **instructional standards** that educators are required to teach. Standards are similar to a 'to-do' list that a teacher can use to guide instruction. Student achievement will increase when quality instruction is used to teach instructional standards.For instance, you have a to-do list that involves three tasks: dropping off the cleaning, filling your gas tank, and studying for a final. Questions you may ask yourself are:

In what order do I accomplish my tasks? How am I going to get each task finished? Should I study at the library where it is quieter or at home where I may be distracted? Is it worth it to purchase gas a few blocks from home at a higher price or drive a short distance to save money? Your goal is to get your to-do list finished in the most efficient and timely way possible.When teaching, you must use the same process when addressing instructional standards. Questions you should ask to successfully complete your 'to-do list' or learning standards in a timely and efficient manner include: What type of students do I have? How am I going to teach the standard? Will they understand the vocabulary? How long do I think it will take for students to fully learn the material?Successful instruction of standards results in student achievement. However, knowing the 'what' and the 'how' is just the first step to successful student achievement. Understanding the factors that can impact a student's ability to learn is equally important.

Factors that Impact Achievement

There are many variables that can impact successful student achievement, but the most critical are classroom instruction and learning disabilities. It is important to remember that all students do not learn the same way or at the same rate. Students are like leaves on a tree; there are no two exactly the same. Just as a leaf comes in unique colors, shapes and sizes, each student has their own unique learning style. we must use a variety of teaching methods and understand the background and individual needs of each student. **Classroom instruction** is the most important factor that impacts student achievement. As a teacher you influence the quality of instruction, set expectations for learning, and measure the level of understanding. For example, when a standard is not presented in a way that a student can understand, or if it's taught in a way that is boring, it can be very difficult for a student to meet the required level of achievement. A good teacher will use strategies such as discussion among students, videos, or stories, to gain student attention and to support the learning process. we should constantly be thinking of ways to make learning fun and appropriate. For example, in looking at our to-do list, you may pre-pay for your cleaning to get a discount or join a friend to make the study session more interesting. Likewise, student achievement involves well-thought out strategies to improve the quality of learning! A **learning disability** is a condition that causes a student to learn at a slower pace than students of the same age or grade level. A learning disability can make understanding of some standards more difficult, but it does not mean a student with this condition cannot achieve academically. It is important to remember that when it comes to student achievement, all students can learn!

DATA COLLECTION

PRELIMINARY NOTES

This chapter is intended to give a detailed description of all the procedures followed by the researcher in order to give the aims of the study. The description includes the following major steps :

1. The population and sample.

2. The instruments and their applications, and

3. The statistical means used for analyzing the computing the result.

4. Population and Sample

5. Population
The sample of study includes of students at Baghdad secondary school who study comprehension. The number is (183) female students, distributed into four sections (A,B,C and D as (45, 46, 47 and 45) students respectively.

THE EXPERIMENTAL DESIGN
The researcher design depended on a random group deign that includes two groups, experimental group which was taught comprehension according to washback technique other control group was taught according to the traditional method. Aspect test has been conducted as shown below :

Table3.1: The experimental Design

Group	Independent Variable	Dependent Variable	Measure of the Variable
Experimental	Backwash	Comprehension achievement	Achievement test
Control	Convention	Comprehension achievement	Achievement Test

3.3 SAMPLE
Since the experimental nature requires two groups one control and experimental, the researcher has chosen classes from second stage in Baghdad Iraqi secondary school girls, English Language lessen, Section A and C being the experimental groups and the other two are the control ones. The number of students in the tow classes are (45) and (46) respectively as shown in tables (2)
All the items have been approved by the expert with some modifications

Table 3.2: Population and the sample of the study

Secondary	School	English language	No. of Students	Total	Sample
Baghdad	Education	Section A	47	183	91
		Section B	45		
		Section C	45		
		Section D	46		

3.4 INSTRUMENTS OF THE STUDY

The nature of this research and its aims requires a balance between instruments :

1.Backwash :

2.Achievement Test on Reading comprehension : Based on Material taught, the researcher has prepared an achievement test of (20) items with four

types of questions, each includes (4) items.

*True and False

*Complete the following....

*Multiple – choice items

*Short- answer questions

Table 3.3: Description of the instrument (testing) or specification of test items

N	Techniques	Items	Degree
1	T.F.	5	5M
2	Cloze.T	5	5M
3	M.C.I.	5	5M
4	S.A.Q.	5	5M

In order to verify the face validity, the test has been subjected to (6) specialists in ELT. In this sense, Ebel (1972:555) states that the best who verify the face validity of the test are the jury members. Hence, all items have been found valid, with some modifications.

5 STATISTICAL ANALYSIS OF THE ACHIEVEMENT TEST (READING COMPREHENSION):

Conducting the statistical analysis of the items is an important step as far as achievement test are concerned. It reveals the ability to what has been intended to measure. Through investigate the difficult coefficient to eliminate the difficult ones and the easy one, and to determine the maximum and discrimination power degree and exclude any other ones. Anastasia indicates that the item should be valid through correlation with internal or the test items, as sample has been chosen randomly, made up of (91) female students classes atAlmudather Secondary school. The total number of other classes was (92) female students after eliminating. After apply the experiment on the sample and answer and calculating the marks according to the psychometric features of the items.
Construction

When teacher constructs his own test is important to make advantage of the experience of other. He should try become familiar with the do's and don'ts accumulated by educator who have designed tests over the years. When he chooses which objectives to measures and what types of items to u se, he should know something about the alternative he has rejected so that he can justify his measurement plan. The teacher should build the test from existing objectives, items, and even subtests-particularly if he has evidence of their quality (Morriseand Fitz-Gibbon, 1978:70). Gronland (1965:109-18) states the following principles of classroom testing:
1.Test construction procedures must take into account the use to be served by the test.

2.The types of test items used should be determined by the specific learning outcomes to be measured.

3.Test items should be based on a representative sample of the course content and the specific learning out comes to be measured, i.e., in accordance with the table of specifications.

4.Test items should be of the proper level of difficulty.

5.Test items should be so constructed that extraneous factor do not prevent the learner from responding.

6.Test items should be so constructed that the learners obtain the correct answer only if he has attained the desired learning outcome.

7.The test should be so constructed that it contributes to improved teaching learning practices.

Test validity

It deals with some aspects of test validity which support the connection between test validly and backwash especially considering the issue developed later in the paper.Alderson and Wlall (1993 as cited in Falcher and Davidson, 2007:223) claim that backwash cannot be related directly to a test validity and Critze the statement of some writers that a test's validity should be measured by the degree to which it has had a beneficial influence on teaching.

Alderson and Wall reject the concepts "backwash validity" this form of validity has never been demonstrated, or empirically rather than asserted (ibid). Messick (1996:3) emphasizes two elements of test properties, authenticity and directness, because they are likely to produce backwash. He classifies both properties under construct validity – Looking at the broader concept of "Validity fame work, backwash is seen as an instance of the consequential aspect of construct validity (1996:242) to encourage positive and reduce negative backwash, test should minimize construct under representation and construct irradiance in the assessment. According to Messick backwash is not simply good or bad teaching or learning practice that might occur with or without

the test, but rather good or bad practice that is evidentially linked to the introduction and use of the test. "If a test validity is compromised because of representation or construct irrelevant Variance it is likely that any signs of good teaching or learning associated with the use of the test are only circumstantial and more likely due to good educational practice regardless of test use similarly, signs of poor teaching or learning associated with the use of a construct validated test are more likely to reflect poor educational practice regardless of test use. Al though there many be exceptions requiring careful scruting, negative backwash with the introduction and use of more valid test because construct under variance bad educational practices while minimizing these threats to validity should facilitate good educational practices (Messick, 1996:247).

Positive backwash is, according to Messick linked to authentic and direct assessments and to the need to minize construct under representation and construct irrelevance in the test.Hughes (2003-7) agrees that testing implies the testing of performance skills with texts and tasks as authentic as possible "if we test dirtily the skills that we are interested in fostering, then practice for test directly the skills that we are interested in fostering, then practice for the test represents practice in those skills" (2003:54). He is very explicit in promoting direct testing. If we want people to learn to write composition, we should get them to write compositions in the test. If a course objective is that students should be able to read scientific articles, then we should get them to do that in the test (2003:54). Two types of validity are determined: content and face validity. The procedures used are discussed below.

Content Validity

Logically, content validity must be evaluated and ensured before face validity. Content Validity refers to extent to which a test measures a representative sample of subject-matter content and the behavioral changes under consideration. It pertains to how well the content of a test measures what is intended to be measured (Grohlund. 1976:Bergaman, 1981). Lado (1961:343) states that in foreign language testing, there is no substitute for content analysis: Statistical study performance of items helps to refine and improve content and performance. Al-kubaisi (1985:11) believe that" content validity is ensured far as the abilities or the element intended to be moniter content validity of the test items. These criteria are:

1. The test item affords a suitable situation (this is decided through type content of the item).
2. It is accurately worded.
3. It contains appropriate vocabulary load.
4. It does not rely heavily on were memeory.
5. It relies heavily on linguistic rather than general intelligence factors.
6. It demands one thing at a time
7. It demand items is preceded by clear instructions.
8. The items do not involve "eye-trick".
9. The test item involves some restatement) rather than were coming).
10. The items must be limited to the material included in the passage.

According to Lado (1961:238) "the items testing comprehension of a reading passage may deal with the total passage, with a particular sentence or part of a sentence in the context of the passage, or with individual words or even parts of words".

Face Validity.

Face validity refers to whether the test measures the objectives behind it or not (Gronlund, 1976) Thorndik and Hagen,

1977: Bergman, 1981). To determine content and face validity reference to the two criteria:

1. Extent of interest to the students (testers)
2. Appropriation of the standard of difficulty, determine the suitability of the test items for testing reading comprehension.

DATA ANALYSIS

This study is conducted to explore the effect of backwash technique on student's achievement in reading comprehension. This chapter is mainly concerned with the presentation and discussion of the result according to the statistical manipulation of the data obtained through the administration of the test to the study subject.

ANALYSIS OF RESULTS;

The obtained results will be presented according to the aforementioned aims of the present study as shown below.

PRESENTATION OF THE RESULTS

The standards means and deviations of the scores of reading comprehension test for the two experimental and control groups.

1. Experimental group : After experimental group, individuals completed their test of comprehensive reading and the analysis of answers according to the type of questions (true/false, fill in blank, multiple choices, and short answers). The arithmetic means and standard deviation of the experimental group scores was in table (5).

To know the significant of the difference of the test of 10 marks for each type of question of 2.5 scores, the researcher used test for one sample (Glass Stanley, 1970). The results are as in table (5).

The arithmetic scores table (5) of the experimental group scores in reading comprehension according to the type of

® International Journal of Research

Available at https://edupediapublications.org/journals

p-ISSN: 2348-6848
e-ISSN: 2348-795X
Volume 04 Issue 08
July 2017

question used and the value of the difference.

Table 4.1: It was found that the experimental group

LEVEL	T		S.D	X	N	Test and Question type
	Table	Calculated				
0551 for the experimental group	3.533	8.1211	3.194	13.867	45	Test (Total
---	--	5.863	1.078	3.444	--	True and False
---	---	5.937	1.270	3.622	--	Complete
	---	5.261	1.258	3.489	--	Multiple Choice
---	---	4.335	1.222	3.289	---	Short Answers

From the table (5) it was found that the experimental group means is greater than the theoretical means of the tests and types of question.

2. Control group : after the control group individuals completed their test of comprehensive reading and the analysis of answers according to the type of questions (True/False, fill in the blank, Multiple choice, short answers). The arithmetic means and deviation of the experimental group scores was in table (6).

To know the significant of the difference of the test of 10 marks for each type of question of 2.5 scores, the researcher used test for one sample. The results is as in table (6).

TABLE4.2: The arithmetic scores of the control group scores in reading comprehension according to the type of question used and the value of the difference.

Significance Level	T		S.D	X	N	Test and Question type
	Table	Calculated				
051 for the experimental group	2.693	3.076	2.778	11.261	216	Test (Total
Not Significant	2.016	0478	0924	2.565	--	True and False
0001 for the control	0.530	3.841	1.115	3.130	--	Complete
001 for the control	2.693	2.909	1.268	3.094	--	Multiple Choice
---	654	4.335	899	2.587	---	Short Answers

Complete the experiment and test for the purpose of this study. To correct the answers and to calculate the scores. It turned out degree with standard deviation of (3.773) while the average scores of the control group (11.433) with standard deviation of (3.263). to know the significance o6f the difference between the two means, the researcher used test for two independent samples for independent. It turned out that the

difference between the two is (4.156) is greater than the table value (3.418). the difference was for the experimental group as the average of the degrees is bigger than the table value. That indicates what we can say that backwash technique helps to contribute to elevate the comprehension for third stage students in English.

Table 4.3: Value of the difference significance for the two experimental and control groups in the scores of comprehension

Significance Level	T		S.D	X	N	Test and Question type
	Table	Calculated				
Experimental	45	13.867	10.202	4.156	3.418	0.001 for the experimental group
Control	45	11.261	7.171			

The relationship between the backwash and the scores of the acquisition in the reading comprehension. The researcher used Pearson correlative coefficient. Thus, the correlative coefficient was (0.385) which is statistically indicated at (01=(0). If the value of the correlative coefficient. This result indicates that the degree of comprehension has a positive relation with student's scores in backwash test (Glass, Stanley, 1970).From the table 4.2 it was found that the control group means is greater than the theoretical means of the tests and the types of questions.

From the table 4.3 it was found the comprehension and distinction item difficulty coefficient the acquisition test for reading

Table 4.4: coefficient the acquisition test for reading

Validity Coefficient	Destination Coefficient	Difficult coefficient	Item
.50	.68	55	1
.33	70	52	2
.45	75	60	3
.45	43	68	4
.61	48	43	5
.27	40	37	6
.38	85	50	7
.52	35	72	8
.31	42	39	9
.57	38	40	10
.57	38	40	11
.38	69	62	12
.20	47	70	13
.39	48	38	14
.47	65	44	15
.42	73	58	16
.39	75	51	17
.42	61	53	18

| .27 | 48 | 55 | 19 |
| .39 | 83 | 40 | 20 |

CONCLUSION, RECOMMENDATION

1)   A study may be carried out to explore the relationship between teachers and   perceptions of washback in other stage (primary, intermediate, and university stage).

2)   A study may be conducted to investigate impact on the part of learners.

3)   A comparative study of the teachers and learners perceptions of backwash impact   is needed.

4)   A study may be carried out for the purpose of promoting positive washbackand   preventing negative washback.

RECOMMENDATIONS

In the light of the results and conclusions of this study, the researcher forwards the following recommendations:

1. Iraqi Secondary school pupils are asked to help their learners in recognize positive washback and using it to increase their achievement.

2. Secondary school student need to be informed with the concept and the use of washback in language and learning.

3. Secondary school pupils should have training courses benefiting from positive washback and avoiding one.

4. Secondary school pupils should be training on the importance of positive backwash and the disadvantages of negative backwash.

REFERENCE:

1. Abraham, paul. (2002). "Skilled Reading Top-down, Bottom-up".Field Ntes.Vol. 10(2).
2. Afflerbach P., P. David person and scott G. Paris (2008) "Clarifying Differences Between Reading Skills and Reading Strategies" The Reading Teaching (15), pp. 364-373. San Franciso: San Francisco state university.
3. Alderson, D. Wall, (1993) "Does backwash Exist? Applied linguistics, Vol. 14pp.115 (Internet source).
4. Alderson, J (1979) The Cloze Procedure and Proficiency in EFL; TESOL Quarterly, Vol. 13, No : 2:219-227.
5. Alderson, J.C (1986) Innovations in M. Portal (ED). Innovations in Language Testing.MEERNELSON, PHILADELPHIA, 93-105.
6. Alderson, J.Charles, (2000) Assessing Reading, London: Longman Group LTD
7. Alderson, J.C. and L. Hamp-Lyons, (1996) TOFEL Preparation courses: a study of backwash. Language Testing Vol.3,pp.280-97.
8. Alderson, J.C., and D. Wall, (1993) "Does backwash Exist?Applied linguistics, Vol. 14pp.115 (Internet source).
9. Al-Jamal, Dina and Ghadi, Nedal. (2007) Englsih language General Secondary Certificate Examination backwash Mu'tahuniversity.
10. Al-Kubaisi, Amir B. "The use of Multiple Choice Technique in testing Reading Comprehension: Discrimination and student's

performance", the educational list, Nos. land, college of Education university of Baghdad, 1985.

11. Almasi,J., (2003). Teaching strategic processes in Reading.Now the Guilford press.

12. Al-Rawi, Salwa A., The Construction of an English Reading Comprehension test for Iraq Secondary School Girls, unpublished M.A. thesis, university of Baghdad, 1972.

13. Al-Saadi, ShathaKathim, An Assessment of Reading Comprehension at university Level, Unpubished M.A. thesis.

14. Andrews, S. (1994). "Backwash or washback? The ` 2c1,Relationship Between Examination Reform and Curriculum Innovation" In D. Nunan, and R. Berry, (eds) Language testing

15. Bachman, B. and A.S. Oalmer, (1996).Language Testing in Practice.Oxford : Oxford university press.

16. Bailey, K.M. (1996). "Working for Backwash : A Review of the backwash Concept in Language Testing" Language testing, Vol-13, No. 3, pp 257-79 (Internet Source).

17. Baker, D. (1989) .Language testing : A critical Survey and practical Guide London: Edward Arnold.

18. Beck, I. (1981), Reading Problems and Instructional Measurement Practice, Research Quarterly, Vol-2, pp:53-95.

19. Bergman, Jerry, (1981) Understanding Educational Measurement and Evaluation.Houghon Mifflin Company, Boston. (Internet Source).