



*Aseorías y Tutorías para la Investigación Científica en la Educación Puig-Salabarría S.C.
José María Pino Suárez 400-2 esq a Lerdo de Tejada, Toluca, Estado de México. 7223898475*

RFC: ATI120618V12

Revista Dilemas Contemporáneos: Educación, Política y Valores.

<http://www.dilemascontemporaneoseduccionpoliticayvalores.com/>

Año: VI

Número: Edición Especial

Artículo no.:22

Período: Agosto, 2019

TÍTULO: Evolución de las pruebas y evaluaciones del idioma inglés: replanteamiento de una cuenta crítica.

AUTORES:

1. Ph.D. Wardah Azhar.
2. Ph.D. Syed Kazim Shah.
3. M.Phil. Muhammad Bilal.

RESUMEN: El estudio investiga la idoneidad de las pruebas de idioma para contribuir no solo a mejorar los procedimientos de evaluación del idioma inglés adoptados en Pakistán por la Junta de Educación Intermedia y Secundaria, Punjab, sino también para investigar su conformidad con cualquier modelo de prueba. Se realiza una revisión histórica bibliográfica con un análisis cualitativo de documentos de preguntas de los últimos cinco años del decimo grado en dicho examen de la junta. Los resultados mostraron que los cuestionarios no cumplían con los objetivos generales del idioma de destino, que se debe prestar más atención a las pruebas de idioma inglés utilizadas en los exámenes de la junta, y que el formato de pruebas debe revisarse para que pueda responder a las demandas de estudiantes y profesores de inglés.

PALABRAS CLAVES: exámenes de idiomas, preguntas, evaluación de idiomas.

TITLE: Evolution of English Language Testing and Assessment: Rethinking a Critical Account.

AUTHORS:

1. Ph.D. Wardah Azhar.
2. Ph.D. Syed Kazim Shah.
3. M.Phil. Muhammad Bilal.

ABSTRACT: This study aims to investigate the suitability of language testing to contribute not only in the improvement of the English language assessment procedures adopted in Pakistan by Board of Intermediate and Secondary Education, Punjab, but also to investigate its conformity with model of testing. A detailed historical account is discussed through literature review, followed by a qualitative analysis of question papers of past five years of 10th class in the said board examination. Results showed that question papers didn't fulfill the general objectives of target language, implying more attention needs to be given to the English language testing utilized in board exams, and that testing format needs to be revised so it can stand up to the demands of English learners and teachers.

KEY WORDS: language testing, question papers, language assessment.

INTRODUCTION.

Instructive educational module over the 'common domains' of the world frequently guarantee to advance impartial and mainstream information through course books and education in formal instructional settings. Be that as it may, investigating these records and the practices fundamentally may enable us to uncover the idea of 'real information'; the political inspiration for such development; and the round of intensity, control and underestimation they include.

Language learning being a standout amongst the most vital pieces of instruction has been abused as a medium through which recursive and coercive practices like 'determination' and 'judgment' are completed successfully. Furthermore, in this way, language strategy of a country is regularly found

supporting the information domain(s) of the power-holding section(s) of the country. It doesn't equitably speak to all areas of the general public as far as the dialects, societies, convictions, political adherence, monetary and economic wellbeing, and so on.

This sort of training may not be obtrusively deliberate. In any case, one may ask why and how such practices discover their way into language arrangements once in a while for the sake of 'secularization' or what Weber calls 'defense'. This is really an exceptionally astute method for developing and embellishment the cognizance of the commanded class(es) without being obviously turning to instruments of social control (Apple, 1979).

Generally seen as an incredible social instrument, language testing is utilized by language approach producers to advance an ideal and favored segment of society as acknowledged ones, and the rest as washouts, disappointments and rejected ones. In spite of the fact that voices have been raised against the set criteria, sustaining nature of tests to make testing learners adaptable haven't been fruitful and tests like IELTS and TOEFL stand as standard testing.

The greater part of the work in language testing hypothesis centers on two zones - the meanings of knowing a language and the fitting strategies for speculating it. In this manner, language analyzers have invested much time and exertion in characterizing the development of language information, as per Spolsky (1968) "Principal to the arrangement of substantial trial of language capability is the hypothetical inquiry of what is known to a language" (p.79). An inimitable definition and distinguishing proof of the structure of language empower analyzers of language; to configure testing methodology which will match such depictions as these will have direct results on the constructive legitimacy of language tests.

DEVELOPMENT.

Framing the Issue.

Appraisal of language is positively as ancient as language tutoring. From the straightforward and single word trial of bygone eras, to perplexing and far reaching test of informative language aptitude of ongoing occasions (Bachman, 1990), and to learning-focused evaluation of the current time (Purpura, 2015), testing of language has experienced exceptional alterations over the span of decades. A large portion of the progressions have been because of the way that language testing has been resting at the intersection of various orders established in connected phonetics, psychometrics (Farhady, 2005).

Connected etymology has been contributory in characterizing, filtering, and reclassifying the idea of the linguistic capacity build; psychometrics has molded the systems for estimating and development that has managed the elucidation of the result of estimating the build in instructive settings. In this way, it appears to be normal to see numerous adjustments in testing language since variations in any of these orders have regularly prompted changes in the elements of testing language.

Making the Case.

Existence of testing language has been around for a considerable length of time without holding to a specific hypothesis amid what Spolsky labeled as the "prescientific" or "instinctive" period.

Language testing is concerned about the estimation of language information. Language information is the 'characteristic' and how we approach estimating it is the 'technique'. Attribute includes the 'what', i.e., the space of language information, and strategy includes the 'how', the suitable systems for estimating language learning. It is the multifaceted nature of the language characteristic that marks a necessity for an exceptional control called language testing; for there is still no full comprehension of what is associated with knowing a language.

In building language tests, it is fundamental thusly to have categorized educational modules or set collection of learning from which analyzers figure out what to test. In the meantime, it is essential to apply fitting psychometric criteria to guarantee that tests developed from such definitions are test-verification, that is, dependable and substantial. The vast majority of the work in language testing hypothesis centers round these two zones - the meanings of knowing a language and the fitting methods for estimating it.

In this way, language analyzers have committed much time and exertion to characterizing the build of language information, as indicated by Spolsky (1968) "Central to the arrangement of legitimate trial of language capability is the hypothetical inquiry of what is knowing a language" (p.79). A reasonable definition and recognizable proof of the structure of language empower language analyzers to configuration testing methodology that will match such portrayals as these will have direct results on the develop legitimacy of language tests.

Prescientific Era.

In this period, language instructors utilized some instinctive strategies following the exercises of what was later called the "syntax interpretation" strategy for teaching. The tests in this period did not compare to standards of any hypothetical system in light of the fact that there was just no hypothesis of instructing or taking exam at the time. Nor were psychometric standards, for example, dependability and legitimacy much being used on the grounds that these ideas had not yet been executed in the field.

Tests were regularly made based on instructors' instinct and made a decision by them as legitimate in light of the fact that tests for the most part included interpretations from or into the objective language that compared to what was really drilled in educating settings.

Scientific Era.

With the conversion of study of language training as a scientific method or process, in the late 1940s and mid '50s, current language instructing started to appreciate the rules offered by logical teaches, for example, semantics and behavioral studies.

Phonetics endeavored to depict the idea of language, and psychology started to clarify the idea of learning forms. The impact of phonetics and psychology in the field of linguistic training marks the start of the logical period in language examination and educating. From that point onward, the greater part of the advancements in language testing have been established in the improvements in the speculations of semantics, psychology, and different overlapped fields such as phonetics, for example, sociolinguistics, second language procurement, and others.

Statement of the problem.

Learning to take examination is vital for excelling in academia. Practically, test-takers of various capability levels may vary in its capacity to relate the given indications in added non-dynamic evaluation setting. With the end goal to thoroughly analyze this capacity and discover which group of test-takers took the best preferred standpoint of the intervention thus, following research questions were introduced:

1. How the level of difficulties can be assessed among the learners?
2. Which skills are involved in testing?
3. Which theories are involved in making a test?

The following examination question aimed at exploration of a distinction in potential dimension of test-takers in applying the clues in non-dynamic evaluation periods after specific time interim. This is to see if members of level of high-, mid-, and low-capability contrasted in their capacity to utilize the focuses amid intervention in non-dynamic sessions later on and if the appropriate response resulted in clear distinction of learner groups as progressively competent in reading skills or less.

Stages in evolution.

English language appraisal has a history that runs parallel to that of English language instructing, and therefore, it appears to have advanced in accordance with changes in way to deal with language teaching. First Spolsky (1975), and afterward Brown (1996), Bachman (2000), and so forth have endeavored to take a gander at the advancements in the field of language appraisal through stages however every one of them concurred that the divisions ought not to be deciphered as watertight compartments with select highlights. Aside from utilizing the real patterns portrayed by the previously mentioned creators, I have likewise included an area the ongoing advances in language evaluation that incorporates the ascent in mindfulness about social measurements and investigation of elective types of appraisal.

The Beginning: English Language Testing and the British Stronghold.

English language testing started in the fifteenth century, when English Language Teaching was in its earliest stages.

Henry V began an English Language strategy as indicated by which French was to be supplanted by English as the language of imperial correspondence. There emerged the need of instructing English to individuals. In any case, the choices identified with educating and testing strategies were taken by coaches. That did not prompt development in educating and testing since it was not permitted to spread among the mass. In any case, after sixteenth century when endeavors were made to characterize and conceptualize language, genuine consideration was paid to creating strategies for instructing English.

With Johann Christian Fick's 'Viable English Course' (1793) and John Miller's 'The Tutor' (1797), ELT was on the track of beneficial research, hypothesis and experimentation. In any case, English language testing needed to hold up until 1913 to take the state of present-day government

sanctioned tests. The entire world does not appear to have broken free of the sorted out separating rehearsed through language tests created by the US and the UK.

Grammar-Translation Approach and Testing.

The pre-Lado period, i.e., the period before 1960s in testing bears a solid impact of Grammar-Translation Approach to language teaching. It is amazing that the recently developing semantic mindfulness in language instructional method did not appear to have influenced language testing in the primary portion of the twentieth century.

The accentuation on point by point investigation of syntax standards and utilizing this information for making an interpretation of sentences from L1 to English (as an L2 or an unknown dialect) and the other way around, kept language figuring out how to retain principles and vocabulary. Be that as it may, the tests were anything but difficult to plan, and doling out imprints was very efficient. A generally amazing case of such tests could be The Charter's Diagnostic Language Test and the Pressey English Test which tried Grammar, Punctuation, Capitalization, and Sentence Structure. Absence of objectivity and measurable examination damaged the proficiency of these tests. Subsequently, this period is called 'pre-logical' by Spolsky (1978).

Structuralism.

In the mid-1960s, the strength of structuralism in semantics and behaviorism in the psychology of learning prompted the rise of the most prominent technique for language instructing alluded to as the "audiolingual strategy".

Following the standards of this technique, Lado (1961) offered the first etymologically arranged hypothetical structure of language capacity. As indicated by his model, language comprised of sounds, words, and sentences showed in the four language aptitudes. Further, language capacity was thought to be the total of the learning of a person. As a pioneer, Carroll (1961) proposed a key

change to language testing and expressed that language tests should endeavor to quantify language aptitudes and segments in an incorporated way as opposed to separating them into little discrete point components.

The Structuralist Approach to Testing.

The Grammar-Translation approach to testing was found inappropriate and ineffective and replaced by what Spolsky (1978) calls a 'psychometric-structuralist' trend in the 1960s. This trend bore the influence of behaviourists like Skinner and structural linguists like Fries and Bloomfield.

Learning of Language was perceived as a method of habit formation, and language testing was a process of measuring language skills and elements of language at discrete levels. Once again, individual student and context were neglected in the name of science, objectivity, validity, reliability and precision. Foucault (1971) calls "appearance of a new modality of power in which each individual receives as his status his own individuality", and in this framework individuals were confined to being 'cases' (Foucault, 1971). The students were the worst sufferers in this case because they were forced to confirm to a set of predicted behavior. This, in turn, resulted in the suppression and loss of natural learning abilities of students.

Integrative Sociolinguistic Era.

Carroll's recommendation set off another time in the field of testing language and was advanced by Oller during the 1960s and '70s as integrative approach and was alluded to as the "integrative sociolinguistic period" as termed by Spolsky (1978).

Holding fast to Gestalt psychological research that the entire is unique in relation to the total of the parts, Oller's principle contention was that the whole of the scores on discrete point things would not mean a sensible sign of test takers' general language capacity. He further trusted that the commitment of discrete direct things of language toward the general language capacity is neither

unmistakably recognizable nor huge. Accordingly, he theorized that language capacity is unitary in nature and all language segments and aptitudes are distinctive indications of this capacity. Rather than discrete point tests, at that point, the improvement and utilization of integrative tests, for example, cloze and transcription ended up well known testing strategies (Oller, 1983).

Oller's purported "unitary factor speculation" produced broad research in the field; for example, Bachman and Palmer (1982) connected multi-trait-multi-method investigation, and Farhady (1983) used distinctive renditions of factor examination and both thought of discoveries unique in relation to those revealed by Oller and his supporters. The result of the reanalysis of a few arrangements of information persuaded that the unitary speculation was chiefly the result of the ancient rarity of off base application and error of the discoveries of factor explanatory strategies. As per these scientists, language capacity comprised of various fundamental factors however a solitary capacity was observed to be basic to every psychological test including language tests.

Integrative Approach to Testing.

The opposition to discrete-point testing gave rise to an integrative approach to language testing, i.e., a combination of psycholinguistic and sociolinguistic approaches. Influenced by cognitive psychology and learning theories, the psycholinguistic view of learning challenged the theories proposed by Saussure and Bloomfield. Chomsky's theories about language and learning acted as a driving force. Language testing, under the influence of this theory, shifted its emphasis from linguistic accuracy to functional ability. Language tests adhered to problem-solving approaches and were expected to reveal what underlying rules the learners had internalized. The sociolinguistic views of language, on the other hand, were guided by Hymes (1972) who emphasized that the social context of a message is as important as its linguistic context. Accordingly, a language test was

expected to measure one's ability to use linguistic elements and skills communicatively and appropriately in a given social situation.

Communicative Era.

At the pinnacle of contentions on the distinctness or unbreakable quality speculation in regards to the idea of language capacity, the field of language testing saw another leap forward with the rise of the informative time in the late 1970s. Canale and Swain (1980) started this development by offering another hypothetical structure for language instructing and testing.

A huge advancement in this time was that etymological skill, which was the focal point of characterizing and estimating language capacity, was viewed as deficient for characterizing and inadequate for estimating language capacity. They asserted that a language capacity structure ought to incorporate parts of correspondence essential in the genuine setting of utilizing language. As indicated by this structure, language capacity or informative skill, as they stated, comprised of three parts including etymological capability that represented what both Lado and Oller viewed as: (a) language capacity, (b) social or talk fitness, and (c) vital ability. Truth be told, notwithstanding the phonetically arranged structure of Lado and the psycholinguistic system of Oller, talk skill and key fitness were added to the meaning of language capacity.

This model was a noteworthy progression in conceptualizing the idea of the language capacity develops and has been very powerful in shaping later systems. Practically all hypothetical and operational medications of language capacity that were presented later had a hint of Canale and Swain's model; for example, Farhady (1983) endeavored to improve the model by expressing that it was direct in structure and added substance in nature implying that open skill would be the total of different capabilities. He additionally contended that informative skill is so huge in area and complex in nature that even local speakers may not accomplish every one of its measurements. He

proposed an intuitive model of language capacity and considered it the "model of practical skill" after the utilitarian notional technique for instructing.

At the point when the multicomponent idea of language capacity, overlooking the quantity of parts, was genuinely settled, a progressively detailed type of Canale and Swain's model was presented by Bachman 10 years after the fact in 1990. His model was extended to incorporate numerous different parts notwithstanding the first three. This model incorporates authoritative information that covers syntactic and printed learning, and logical learning that covers practical and sociolinguistic learning. Bachman's model is unique in relation to others in degree as well as in hidden suppositions, in that Canale and Swain's model appears to manage open skill as predominantly the capacity of language students to impart in reality, though Bachman's model speaks to language capacity managing correspondence in numerous different settings including the scholastic setting. That is the reason extra kinds of learning, for example, literary, and utilitarian are incorporated into this model.

The previously mentioned hypothetical systems offered to clarify the build of language capacity are very uncovering, particularly when considered sequentially. Each model is fairly more complete than the first ones, showing language capacity in an increasingly mind-boggling way and in the meantime progressively hard to gauge.

Communicative Approach to Testing.

Hymes' model of "communicative competence" continued to guide the field of language testing in Europe and America till the 1980s. In 1980s and the early 1990s, the language testing models proposed by Morrow (1979), Canale and Swain (1980) and Bachman (1990) who emphasized on testing both, competence and performance of the learner gained in popularity and importance. Bachman (1990) defined language ability as a combination of two components: "language

competence”, i.e., a variety of language knowledge and “strategic competence”, i.e., a set of metacognitive strategies.

Bachman and Palmer (1996) took up from where Hymes had left. They argued that the construct and context of tests must be defined clearly; the materials and test tasks must be as authentic as possible; and real-life situation must form the background of all test items. In addition, they asserted that a test must take into account and measure linguistic, sociolinguistic, discourse, and strategic competences.

Discrete Point Testing.

In the discrete point time, language tests concentrated fundamentally on the appraisal of separated and discrete point things, for the most part syntactic and lexical, and following auxiliary etymological standards of the time. Consequently, tests in that period incorporated various decision, genuine false, and different kinds of target things which concentrated on single and free things like conjugation of action words and distinguishing proof of lexical components in a decontextualized way. These tests were connected to all language aptitudes - perusing, composing, tuning in and talking, and evaluated discrete and segregated parts of language. Indeed, even a profitable ability, for example, composing was tried such that required test takers to recognize distractors which contained mistakes through different decisions testing, as opposed to asking test takers to deliver genuine composed language tests.

Testing Integrative Language.

In the integrative period, language tests were seen in a comprehensive and more contextualized way concentrating on the testing of worldwide language tests - complete passages and full messages. Testing errands included composition letters and appreciation of entire writings with insignificant reference to separated components in the content. In this period extraordinary consideration was

given to a particular kind of tests, the cloze, in which words were erased from longer messages and the test taker was relied upon to fill in the missing openings.

Oller (1975, 1979) who advanced the cloze, guaranteed that it tapped integrative language and mirrored a unitary thought of language which underlies the language information dependent on the student's sober minded sentence structure of anticipation. He battled that this information spoke to a mental portrayal of the language client's capacity to delineate onto settings. Oller contended that the sentence structure of anticipation was the central instrument basic the abilities of reasoning, getting, talking, perusing and composing, and it was to be enacted in conditions which required the handling of language under typical relevant limitations. This capacity was to be operationalized through integrative tests, for example, the cloze (and transcription) since in these tests, students needed to prepare their etymological and additional phonetic learning to reconstitute the importance of content.

Critical Language Assessment and Alternatives in/ to Assessment.

Language testing has been growingly perceived at as a political and ideological wonder profoundly established in the public arena and culture. All the while, endeavors have been made to investigate these elements of language testing. The way toward taking a gander at testing began long time prior when Henry Latham (1877) scrutinized 'infringing power' of examinations which he thought had a biasing impact on training. Almost a century later, similar kind of complaint was heard from Foucault (1977): "...the examination is at the center of the procedures that constitute the individual as effect and object of power, as effect and object of knowledge. It is the examination which, by combining hierarchical surveillance and normalizing judgement, assures the great disciplinary functions of distribution and classification, maximum extraction of forces and time, continuous

genetic accumulation, optimum combination of aptitudes and, thereby, the fabrication of cellular, organic, genetic and combinatory individuality”.

Foucault might have been inspired by Paulo Freire’s (1970) masterpiece ‘Pedagogy of the Oppressed’. But both of them surely had some impact on the discipline of Applied Linguistics which was taking shape during the last part of 1970s. However, language education, especially language testing took a lot of time to accommodate critical perspectives in its discourse. The concepts of *anxiety, bias, hegemony, democracy, marginalization, dominance, ideology*, etc., started to be taken seriously only in the 1980s, i.e., the time when Norman Fairclough (1989) published his monumental text ‘Language and Power’. This again shows that educational changes can be felt more strongly in the form of political visions than practical options.

During the 1990s, experts in language testing like Spolsky, Tim McNamara, Elana Shohamy, etc tried to go beyond the linguistic boundaries and look at language testing as a powerful educational tool that is used for social, political, cultural, and above all, ideological control. These advocates of fairness and ethics in language testing got solid support from the evolving discipline of Critical Applied Linguistics. The result is evident in concepts like *critical language testing, democratic assessment* (Shohamy, 2001), *alternative assessments* (Huerta-Macías, 1995), *fairness* (Kunnan, 2000), *bias* (Shepard, 1981), etc.

Looking critically at language testing has led to search for alternatives ‘in’ and ‘to’ assessment. Brown and Hudson (1998) list checklists, journals, logs, videotapes, audiotapes, self-evaluation, teacher observations, portfolios, conferences, diaries, self-assessments, and peer assessments as alternatives in assessments. These alternatives may be used as substitutes *to* what we call ‘testing’. They provide options that may help us move beyond technicalities of language testing and endear testing as a useful tool that can promote learning in formal educational centers. Moreover, they have

paved way for a broad framework that can accommodate a variety of individual learning styles and preferences while acknowledging the identity and abilities of every single learner.

Dynamic Assessment.

Dynamic appraisal in language learning, was introduced by Vygotsky's (1986, 1978) thought on how children's cognizance creates and applies Vygotsky's sociocultural hypothesis into evaluation, can offer new experiences into appraisal in the language classroom by uncovering priceless privileged insights about the capacity of individual learners and their capacities while noting each test item. The reason can be the procedure which employed the nature of dynamic appraisal. While the consequences of non-dynamic appraisal can just demonstrate the officially existent capacities of the learner, the examination of ZPD makes it conceivable to assess the capacity of the learner to gain from the communication with an educator or a more skilled friend.

To underscore the liquid idea of dynamic evaluation, Lidz (1987, p.4) characterizes it as "an association between an analyst as-intervener and a student as-dynamic member, which looks to gauge the level of modifiability of the student and the methods by which positive changes in intellectual working can be actuated and kept up". According to Lantolf and Thorne (2006, p.28), Vygotsky contended that "The main suitable method for comprehension and clarifying ... types of human mental working is by concentrating on the procedure, and not on the result of improvement". This is the basic point which recognizes dynamic evaluation from non-dynamic appraisal. Murphy (2011, p.1) considers DA to be "a way to deal with comprehension and considering a person in the appraisal procedure".

DA gives critical data to successful remediation, which is a definitive objective of this evaluation and isn't given by conventional non-dynamic tests. Lidz (1995) saw that conventional institutionalized evaluation trails the student's psychological advancement to the point of

"disappointment" in his/her autonomous working, though DA drives the learner to the point of making progress in intervened execution since it goes for distinguishing deterrents to more viable inclination and execution, to discover approaches to conquer those hindrances on ensuing learning and execution viability in the research of Haywood and Lidz (2007, p.3).

The presumption behind powerful appraisal is that a few people can accomplish substantially more intellectually whenever furnished the chance to work with a 'critical other' to enhance their psychological proficiency. The point of dynamic appraisal is to streamline psychological working, instead of just to test it, and it is here that a change in outlook in scholarly evaluation happens such as pointed out in the works of Grigorenko and Sternberg, (1998, p. 77); Lidz, (1997, p. 291). An imperative preferred standpoint of DA is making suggestions dependent on formative potential which isn't uncovered by customary non-dynamic tests (Davin, 2011).

In powerful appraisal, the students are told on the most proficient method to play out specific undertakings, and intervened help on the best way to ace them are given. Their advancement in the capacity to take care of comparative issues is then estimated by Kirchenbaum (1998).

Lidz (1987) sees dynamic appraisal as a connection between an examiner as-intervener and a student as-dynamic member, which looks to gauge the level of modifiability of the student and the methods by which positive changes in subjective working can be initiated and kept up. He characterizes dynamic evaluation as: "...ways to deal with the improvement of choice explicit data that most typically include association between the analyst and examinee, center on student metacognitive procedures and responsiveness to intercession, and pursue a pre-test– intervention– post-test regulatory organization" (1997, p. 281).

Population and Sample.

Research is invariably conducted by means of a sample drawn from the target population on the basis of which generalizations are drawn and made applicable to the population. There are, in all, 10 Boards of Education/Examination conducting public examinations in the Province of Punjab. In fact, every district division has a Board of Education/Examination.

The target population in the present study covered however comprises only the Faisalabad Board of Intermediate and Secondary Education/Examination. The selection of the Board was made on the basis of convenience for the researcher and past 5-year papers were selected as a sample to create relevance with the current practices. The simple selection of the board was made for obtaining the design, blueprint and the marking scheme.

Methodology.

In the present investigation, the procedures created by NCERT for exam tests paper, for example, Design, blueprint, stamping plan and question shrewd examination and so on., were utilized. Despite the fact that the Faisalabad board utilize the above said strategies in paper setting for exam of matriculation, ordinarily it has been experimentally discovered that the inquiries set by the Board are not in accordance with the dimension of instructional targets as illustrated in the plan and blue print. Consequently, it isn't just exacting adherence to the above mentioned, yet in addition to check the nature of the tests in order to discover how far they satisfy the guidelines is significant. This needs a subjective examination of inquiry papers notwithstanding the quantitative systems of investigation through the portion of imprints to various types of inquiries and time limit and so forth.

Procedure of data collection.

For this study, the Question papers of matriculation level ten from past five years were selected from Faisalabad B.I.S.E inclusive of their design, blueprint, marking scheme and question wise analysis.

Analysis and Interpretation.

After the accumulation of information, the information was classified, and rates were determined. A subjective examination was made based on the acquired information as far as the Bloom's Taxonomy of Instructional Objectives to check regarding how properly they fit into the endorsed plan on various parts of inquiry papers as pursues in Table 1:

A detailed qualitative analysis discussed later will reveal more insight into the implications of this allocation.

Sr. no.	Year of Paper	Content	Percentage as calculated from total marks of the paper
1	2018	Grammar	38.7%
		Translation	21.3%
		Essay type question	40%
2	2017	Grammar	38.7%
		Translation	21.3%
		Essay type question	40%
3	2016	Grammar	38.7%
		Translation	21.3%
		Essay type question	40%
4	2015	Grammar	38.7%
		Translation	21.3%
		Essay type question	40%
5	2014	Grammar	38.7%
		Translation	21.3%
		Essay type question (taken from prose and poetry)	40%

Table No.1 Content-wise distribution of marks.

Table 2 displays information on how the learners may have been given appropriate time for attempting all parts of questions, however, this table does not take into account the parts of each question and how some parts require more time based on their level of complexity and some require less time.

Sr. No.	Marks/Questions/Minutes	2018	2017	2016	2015	2014
1	Total number of questions	8	8	8	8	8
2	Total number of marks	75	75	75	75	75
3	Total number of minutes	150	150	150	150	150
4	Minutes per question	18.75	18.75	18.75	18.75	18.75

Table No.2 Allocation of marks/questions/minutes.

Interpretation of results.

If we see the average amount of time allocated for each question, it has remained constant over the years, and so has the content allocation based on percentage of total marks: however, the distribution of time remains unjustified (will be discussed in detail in the section of qualitative analysis) as not all questions carry equal marks or have equal level of difficulty. A detailed year-wise Question paper analysis is as follows:

Question 1.

This question has 19 subparts, for which to begin with the average time allocated remains only at 18.75 minutes: which is unsuitable for any candidate to attempt to his/her full potential, though this section marks the allocated time at twenty minutes, which still remains unjustified as it leaves no room for reading or understanding any subpart of question for the L2 learner. Moreover, all sub parts are MCQ's in nature, however, the construct of questions are of knowledge learning level and

relatively more in number, which should be diminished. Rather MCQs must be utilized for a wide range of testing (counting comprehension and application level).

The most interesting finding is that all these 19 sub questions of question1 have the allocation of Grammar content; since grammar has to be “applied” to gain language proficiency, the inefficiency of this test is very apparent, as all Grammar based questions are tested only on Knowledge level and that from MCQ type questions with less than a minute allocated for each subpart. (MCQs) are of only knowledge level questions are used.

Question 2.

This question has eight subparts to it; however, students have been given a choice to attempt only five of them, thus, for the analysis only five will be considered here. Each subpart carries only two marks and a total of ten marks.

On an average, each subpart gets only 3.7 minutes of time allocation for two marks. The questions have been taken from the text taught and all of them belong to the Knowledge level (Bloom’s Taxonomy). Out of eight questions, six start with the action verb “what” and the other with “which”, and this usage of these verbs itself makes it clear that they are of knowledge level only. Other two begin with “How”, whereas, at this secondary level, questions from other levels such as, comprehension, analysis must have been included.

The action verbs utilized in the sub-part for two imprints is unseemly in light of the fact that for a two-mark question the hopeful can't legitimize. Henceforth the activity action words of this benevolent must be utilized in article/long answer type questions as opposed to short answer or short answer questions. The model answer given additionally makes reference to just two explicit focuses. Rather a couple of more focuses ought to be included with the goal that the understudies get adaptability in replying. In article type addresses the stamping plan notwithstanding giving

model answer ought to likewise feature/underline explicit central focuses. This will empower the between inspector fluctuation to be as smaller as could be expected under the circumstances.

The guide pointing likewise tests ability and comprehension of the understudies instead of ID and area of spots in a repetition technique. The quantity of target questions could be expanded. Additionally, they must be given in the start of the inquiry paper, as opposed to at last.

Questions 3 and 8.

The question directs the learners to translate the paragraph into Urdu (local language) or to rewrite in simple English while question 8 requires the candidates to translate an Urdu Passage into English or write ten descriptive sentences on the given topic.

The question itself has got a balanced composition of knowledge and understanding level to it. For this type of question, the expected content has to be sufficiently large, but whereas in the case of this question only three sentences are put out there for which the answer cannot be sufficiently long for the award of 8 marks. Moreover, this has to match with the other question given as internal choice.

Question 4 and 5.

This question requires the examinee to write a summary of some poem or to paraphrase the given stanza. And question five directs the examinees to write an essay on one of the three given topics. But, in article type questions the checking plan does not withstand the given model answer ought to likewise feature/underline explicit central focuses. This will empower the in-between analyst changeability to be as smaller as could be expected under the circumstances. The guide pointing additionally tests aptitude and comprehension of the understudies as opposed to distinguishing proof and area of spots in a repetition technique. The paper type question like this inquiry, the competency tried isn't quite certain and has all the earmarks of being exceptionally shallow. It is

neither of analytical in nature, nor involves any reasoning, since all the given topics of essays are of descriptive nature. No argumentative or narrative essay options are given to the learners.

Question 6 and 7.

These questions check the syntactic proficiency of the candidates. Question 6 requires the examinees to change any five of the given eight direct sentences into indirect form. And question 7 elicits the examinee to use any five of the given pair of words into sentences so that the difference in their meanings may get clarified. However, the mark division and time allocation for this task is inappropriate.

Findings of the study.

The study was conducted to analyze the English Question papers of past five years of Board of Intermediate and secondary education Faisalabad. The findings are presented as under:

1. The substance astute circulation of imprints which isn't uniform among inquiries. High significance is given to essay type questions with 40% of questions demanding essay type responses and that too only of descriptive type, which seems inappropriate for the level of students.
2. Different content areas in the test paper give the impression of them being spontaneous and sloppy dissemination of questions. Division of questions with due weightage to every single zone of content variation require to give an equalization, which is basic in building up a coordinated comprehension and capability of language.
3. There is no wide variety in the quantity of inquiries that an understudy needs to reply. Anyway, the expansion in number of inquiries will prompt physical perusing heap of the understudy notwithstanding intermittent mental burden.

4. With respect to the all-out number of imprints and number of minutes/measure of time there is a uniform example for recent years notwithstanding, the assigned time is improper. Despite the fact that number of inquiries themselves does not clarify much, yet it is important to remember that a legitimate dispersion of various types of inquiry paper is fundamental for setting a sound kind of inquiry paper with ideal weightages of time apportioned to various types of inquiries.
5. There is an unpredictable utilization of activity action words which deludes the competitors numerous a period. For instance, the activity action word "legitimize" is utilized for a two-mark short answer type question. This word will be increasingly fitting to use for a long answer type question than a short answer type question. An understudy may, see this activity action word expound his contention to legitimize and all things considered he won't be adequately compensated as the most extreme imprints is just two. This will devour his/her time which generally could have been utilized for addressing different inquiries.
6. An significant factor which the Boards, evaluators and paper setters need to note is that while it is critical to give a perfect answer in the stamping plan to fill in as a guide for granting marks, the evaluators ought not carefully stick to it and license for variety in articulation as far as introduction in language, style and substance and so on. This must anyway rotate around a few central focuses which should basically be incorporated into the appropriate response as generally understudies may compose something disconnected/unessential when they don't have the foggiest idea about the appropriate response. Thus, this basically could be said as adaptability with specific constraints in-worked in addressing questions.

CONCLUSIONS.

This paper is an attempt at revisiting the history of English language testing in critical light. The aim is not to deny the necessity of tests. From the ancient example of the Shibboleth test in the

Bible to the modern day tests like the Australian Dictation Test (1901), the Golden Rule Settlement (1940), Fruit Machine (1950), the Occupational English Test (1983), TOEFL, etc. language testing has proved to be more inhuman than human; more autocratic than democratic; and more a device of control than one promoting freedom. It is high time we mold it in such a way that it accommodates our multilingual and multicultural diversities; and individual preferences, personality traits, ideologies and beliefs.

Suggestions for further study.

1. A similar study can be conducted on large scale i.e., all the boards to ascertain the true situation among different boards of education/examination in Punjab initially and then other provinces of Pakistan, eventually leading to a comparative study at provincial level.
2. The present study was confined to only English of class X of Faisalabad board and similar studies could be conducted in other subject areas.

BIBLIOGRAPHIC REFERENCES.

1. Apple, M. W. (1979). *Ideology and Curriculum*. London: Routledge.
2. Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language Testing*, 17, 1-42.
3. Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford: Oxford University Press.
4. Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford, England: Oxford University Press.
5. Bachman, L. F., & Palmer, A. S. (1982). The construct validation of some components of communicative proficiency. *TESOL Quarterly*, 16, 18–34.

6. Bloom, B.S., ed. (1956). *Taxonomy of Educational Objectives: Handbook I, Cognitive Domain*, New York: David McKay Company, Inc.
7. Brown, J. D. (1996). *Testing in language programs*. Upper Saddle River, NJ: Prentice Hall Regents.
8. Brown, J. D., & Hudson, T. (1998). Alternatives in language assessment. *TESOL Quarterly*, 32(4), 653-675.
9. Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1, 1-47.
10. Carroll, J. B. (1961). Fundamental considerations in testing of English language proficiency of foreign students. In H. B. Allen & R. N. Campbell (Eds.), *Teaching English as a second language: A book of readings* (pp. 313–21). New York, NY: McGraw-Hill.
11. Fairclough, N. (1989). *Language and Power*. London: Longman.
12. Farhady, H. (1983). On the plausibility of the unitary language proficiency factor. In J. W. Oller, Jr. (Ed.), *Issues in language testing research* (pp. 11–28). Rowley, MA: Newbury House.
13. Farhady, H. (2005). Language assessment: A linguametric perspective. *Language Assessment Quarterly*, 2(2), 147–64.
14. Foucault, M. (1977). *Discipline and Punish: The Birth of the Prison*. London: Penguin Books.
15. Frey, B. B., & Schmitt, V. L. (2007). Coming to terms with classroom assessment. *Journal of Advanced Academics*, 18, 402–23.
16. Freire, P. (1970). *Pedagogy of the oppressed*. New York, NY: Herder & Herder.
17. Gronlund E Norman, (1965). *Measurement and Evaluation in Teaching*, The Macmillan Company, New York.
18. Huerta-Macias, A. (1995). Alternative assessment: Responses to commonly asked questions. *TESOL Journal* 5, 8–11.

19. Hymes, D. (1972). On Communicative Competence. In J.B. Pride & J. Holmes (Eds.). *Sociolinguistics*. Harmondsworth, England: Penguin Books.
20. Kunnan, A. J. (Ed.). (2000). *Fairness and validation in language assessment: Selected papers from the 19th language testing research colloquium, Orlando*. New York: Cambridge University Press.
21. Lado, R. (1961). *Language testing: The construction and use of foreign language tests: A teacher's book*. New York, NY: McGraw-Hill.
22. Latham, H. (1877). *On the Action of Examinations Considered as a Means of Selection*. Cambridge: Deighton, Bell.
23. Messick, S. (1995). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13–103). New York, NY: Macmillan.
24. Morrow, K. (1979). Communicative language testing: revolution or evolution? In Brumfit, C.J. and Johnson, K., (Eds.). *The communicative approach to language teaching*. Oxford: Oxford University Press, 143–57.
25. National Curriculum Frame (2005). National council of Educational Research and Training, Sri AurobindoMarg, New Delhi.
26. Oller, J. W., Jr. (Ed.). (1983). *Issues in language testing research*. Rowley, MA: Newbury House.
27. Purpura, J. E. (2015, March). Learning-Oriented assessment in second and foreign language classrooms. Paper presented at LTRC, Toronto, Canada.
28. Sreekanth Y (2004). Societal role in development of creativity among children. *Edutracks*, Vol.4 No.4, Neelkmal Publications (P) Ltd., Koti, Hyderabad, India pp 14-15.

29. Sreekanth, Y. (2005). Environment & Learning', Environmental Education and Awareness, Proceedings of ICEM'05 (International Conference on Environmental Management), BS Publications, 4-4-309, Giriraj Lane, Sultan Bazar, Hyderabad A.P
30. Sreekanth, Y. (2006). Educational Evaluation at School Level: Success and Failures. Edutracks, Vol.6 No.2, Neelkmal Publications (P) Ltd., Koti, Hyderabad, India pp 20-22.
31. Shepard, L.A. (1981). Identifying bias in test items. In B.F. Green (Ed.), New directions in testing and measurement: Issues in testing-Coaching, disclosure, and test bias. San Francisco: Jossey-Bass, 79-104.
32. Shohamy, E. (2001). Democratic assessment as an alternative. Language Testing. P.373 – 392.
33. Spolsky, B. (1978). Introduction: Linguistics and language testers. In B. Spolsky (Ed.), Advances in language testing series: 2. Arlington, VA: Center for Applied Linguistics
34. Spolsky, B. (1978). (Ed.). Approached to Language Testing. Papers in Applied Linguistics.
35. Spolsky, B. (1993). Testing of English of Foreign Students in 1930. ERIC 2002.
36. Spolsky, B. (1975). Language testing – the problem of validation. In L. Palmer & B. Spolsky (Eds.). Papers on Language Testing 1967-1974. Washington, D.FC.: TESOL. 147-153.

DATA OF THE AUTHORS.

1. **Wardah Azhar.** Ph.D Candidate, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan. Email: azharwardah@gmail.com
2. **Syed Kazim Shah.** Assistant Professor in English, Department of Applied Linguistics, Government College University, Faisalabad, Pakistan. He got his Doctoral Degree in Applied Linguistics from International Islamic University, Islamabad, Pakistan. E-mail: kazim.shah@gcuf.edu.pk

3. Muhammad Bilal. Lecturer in English, Department of English, University of Education Lahore,
Faisalabad Campus, Faisalabad, Pakistan.

RECIBIDO: 4 de julio del 2019.

APROBADO: 17 de julio del 2019.