**ASSESSMENT OF GRADE XII IN ENGLISH TESTS WITH EVALUATION OF TEACHER’S INSTRUCTIONAL METHODS IN THE PUBLIC SECTOR 2019-2021**

Baligua Unis

Lahore College for Women University, Lahore, Pakistan

[Baliguaunis@gmail.com](mailto:Baliguaunis@gmail.com)

Naila Alam

Kinnaird College for Women, Lahore, Pakistan,

[naila.alam@kinnaird.edu.pk](mailto:naila.alam@kinnaird.edu.pk),

Fakhra Aziz

Lahore College for Women university, Lahore

[fakhraaziz2@gmail.com](mailto:fakhraaziz2@gmail.com)

**Abstract**

This research study is an objective assessment of public sector Grade XII female learners in compulsory subject, English and their evaluation of teachers’ methodology of instruction following pandemic – related closures in educational institutions and the shift to blended learning. The researcher selected public university Semester 1 students in Lahore. The researcher correlated their performance in conventional learning to the current scenario. Objective tests were administered to students (n=85, n=87, n=88, n=91). National test (alpha 0.547 SEM 0.329) and international test (alpha 0.804 SEM 0.272) are coefficients of equivalence measures. The reliability of the test instrument indicates that it is a good instrument for classroom testing with some difficult items which can be revised. Feedback on Likert scale questionnaires from students indicated the teachers’ way of instruction in blended learning is communicative. Questionnaire reliability (Cronbach’s Alpha 0.865 valid response 60.9%) indicates the Directive teacher spearman’s rho correlation is 1.000 while the Communicative teacher correlation is 0.586 (sig.2-tailed .000 n=100). This research recommends regular revision and objective testing in the compulsory subject, English.

**Key words: *Likert scale, Questionnaire, objective test***

**1. Introduction**

This descriptive research involved Grade XII female Semester 1 students who were assessed in the compulsory subject, English following Covid-19 pandemic closures of educational institutions and the transition to blended learning in Punjab.COVID-19 seriously impacted students, instructors, and educational organizations around the globe (Mailizar, 2020). Online learning is effective in digitally advanced countries (Basilaia, 2020) but may not be appropriate in developing countries. Formative teacher made tests grade students as motivators (Arter, 2010) The Teachers of English to Speakers of Other Languages (TESOL) levels of students’ performance assessment are: Starting, Emerging, Developing, Expanding and Bridging. For Grade XII, the researcher feels Level 3 Developing is a good indication of learners’ ability. Learners comprehend language used in familiar surroundings, use general and academic expressions, and expand oral and written sentences. Learners in assessment may make errors in oral and written language that impede communication but still, retain meaning (Shrum, 2015) Formative teacher made tests grade students as motivators (Arter, 2010)

Students’ feedback on the teachers’ instructional methods in Grade XII helps teachers’ task identity and teachers’ perception of their work as being important, while task significance is how the teacher’s job impacts others. Dialogue in the workplace and classrooms aligns learning objectives with assessment goals (Boyle, 2020). Teachers directly benefit from evaluation (Miller, 1990) while learners project teacher’s self-esteem and self-efficacy (Wolf, 1991). Directive teacher behavior is preferred for speediness at work, silent learners appear competent and knowledgeable, and class work is well-organized. Communicative teachers’ supervision promotes collaborative growth even though communicative teachers may be more idealistic than practical in situations where learners expect firm direction from their supervisors (Doughlas, 1960).

Successful blended learning or E-learning relies on asynchronous learning. Learners benefit from making choices in objective tests and value knowledge from assessment (Cook, 2016). Randomly parallel tests generalizability theory is a domain-sampling approach which assumes two tests to be randomly parallel, with items assumed to be randomly drawn from the same pool of possible items. This research assessed variances across subjects, to differentiate levels of abilities in female students of Grade XII, the objects of measurement. Score variances is the difference in participants score or true variance. Variances of scores between two parallel forms of tests are error variances. True score is a constant. Norm referenced testing establishes reliability of the total score. Content familiarity with the textbook is a prerequisite for reducing test anxiety for learners since the content is a familiar stimulus. Objective testing is a processing theory of deep encoding of content. Stimuli and events actively processed for meaning are remembered (Kosslyn, 2015). Encoding specificity suggests memory recall improves when conditions at encoding match conditions at retrieval. Learning amidst challenging surroundings is associated with being challenged to prove oneself and content is easier to recall in similar testing conditions (Ebbinghaus, 1885 cited in Shrestha,2017) Proficiency in a second language has two different aspects: face to face communication (known as basic or written contextualized language skills) and academic uses of language such as reading and doing grammar exercises (known as academic English) (Garcia, 2002; Snow, 2000)

Smith (1982) in “*A study of Mind, Meaning, and Language”*, has remarked on Bakhtin’s metalanguage and use of the term “heteroglossia” “intonational quotation marks”, and “word-with-a-sidewards-glance” which refers to interaction, as a way of referring to a mode of fixed transcription with repeatable figures referring to the context in which words are spoken since a mix of discourses serve best for mediating their own intentions. Discourse with a mix of intonation, punctuation, lexical choice, and gesture sends out a message. Pre-existing meanings are inherent in dictionaries or ideologies, and the otherness of intentions present is the dialogue. "Reading and writing float on a sea of talk" thus, establishing the primary role of oracy in language development (Fisher, 2008)

Lingard, Hayes, and Mills (2006) noted that in developing countries classrooms with higher number of students enrolled in each class teachers talk more, and students talk less. In present times, diversity in the classroom in the local context implies that the teacher is organized to instruct students of different ages, varied interests, and abilities, from different socio-economic groups and multiple intelligences in a large group handling the strength of large classes, sporadic, irregular assessment submissions, and attendance records amidst the challenge of meeting students online as well as face-to-face. Learners are not a monolithic group and differ in proficiency levels in home language and second language, pronunciation patterns and orthographic representation; hence, they might not be able to recall and relate new concepts to prior learning and adjust to blended learning. Online learning may not be entirely satisfactory since blended learning or E-learning relies on asynchronous learning with access to an electronic network. Furthermore, linguistic flexibility is an asset to learn a new language but for the non -native speakers of English, English remains a challenging subject. Second Language Learners benefit from making choices in objective tests rather than simply relying on being told, and value knowledge about second language acquisition so they can modify the way they work, and make intelligent decisions (Cook, 2016)

**Objectives of Study**

1. What are the effects of blended learning on students’ English tests?

2. How can teachers ensure student satisfaction with blended learning?

**Methodology –** **Research Design Phase 1**

The quantitative descriptive research study was completed in two phases.

**Participants Inclusion Criteria** The researcher compared scores of the high ability, average and low ability regular students of Grade XII.

**Materials and Procedure**

**1 Assessment tools (i)- ENGLISH TESTS**

**Findings:** Phase 1. A local test and an international test were consecutively administered. For local test, the prescribed textbook Grade XII Intermediate English Simple Grammar and Composition (Part II) was used. Encoding specificity suggests memory recall improves when conditions at encoding match conditions at retrieval. Learning amidst challenging surroundings is associated with being challenged to prove oneself and content is easier to recall in similar testing conditions (Ebbinghaus, 1885 cited in Shrestha,2017) The local test included one detailed dialogue writing activity (Khan, 2015). The international online published Nelson Denny Vocabulary Test (Test, 2020) multiple choice items included high frequency words, and sentence correction exercises to assess practice induced improvement. Scores on two tests of 100 marks were compared. Criterion referenced assessment was made with item analysis on each item.

**Table 1 *Parallel Measure Reliability***

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Parallel Measure Reliability | Mean | Standard Deviation | Max. | Min. | Range | Tukey’s Hinges | IQ | Wechsler | Binet | Z scores |
| National Test | 18.6092 | 5.122 | 33 | 8 | 25 | 15, 19, & 22 | 33.96-22.73 | 24.67, 18.92 | 25.26, 18.81 | 3, -1 |
|  |  |  |  |  |  |  |  |  |  |  |
| International Test. | 19.68 | 5.499 | 37 | 8 | 29 | 6, 19 and 23. | 36.126, 22.73. | 22.33-17.03 | 22.85, -17.22 | 2, -2 |
| Secondary School Exam | 751.1163 | 112.22 | 533 | 157.8 | 376 | 699, 788, 842 | 890.4, 889 | 6.430, 5.92 | -5.90- 5.91 | 1, -1 |
| Secondary School English | 0.150 | 29.95 | 170 | 40 | 130 | 65,79, 110 | 150.05, 118.1. | 40.39, 1.330 | 6.816, 0.75. | 2, -2 |

**Table 2** *Reliability estimates of the parallel forms of the test*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Split half Reliability National Test** | **Mean** | **Variance** | **Standard Deviation** | **N of Items** |
| Part 1 | 4.85 | 9.645 | 3.106 | 25 |
| Part 2 | 3.70 | 8.532 | 2.921 | 21 |
|  | .464  .407 | .251  0.244 Common variance  0.072 True variance  0.171 Error variance |  |  |
| **Split half Reliability International Test** | **Mean** | **Variance** | **Standard Deviation** | **N of Items** |
| Part 1 | 4.84 | 11.532 | 3.396 | 25 |
| Part 2 | 3.84 | 2.977 | 1.731 | 25 |
|  | 8.62 | 23.128  2.480 common variance  .098 true variance  2.383 error variance | 4.809 |  |

The Secondary school exam is a congeneric measure. The IQ scores English in Secondary School are 66.87th IQ percentile or 33.124% of the IQ scale. Cronbach’s Alpha for National Test standardized items is *0.520.* Guttman Split Half Coefficient for the national test is 0.461. The reliability of the scale is 0.421 and the reliability unbiased is 0.420. The Test for Goodness of Fit Chi square value is -119.458 degrees of freedom 6 Sig 000. The reliability of scale is 0.896 and the reliability of scale unbiased is 0. 900.The common inter- item correlation is 0.292. The valid number of cases is n=88, 98.9 % and the excluded is 1.1%. The inter item correlation in national tests is 0.28. The Spearman Coefficient equal length and unequal length is 0.468, 0.469. Log of determinant of unconstrained matrix is .000 and the constrained matrix is .34. 69. Between Items F value is 33.895 Sig .000. Cronbach’s Alpha Reliability for International Test is *0.804* based on standardized items with n =87 Cronbach’s Alpha based on standardized items is 0.812 n of items =50. Percentage of test items included 97.8% with 2.2% of items excluded. The n of items=50. Spearman Brown Coefficient Correlation Equal length is 0.845 and unequal length is 0.845. The common inter-item correlation is 0.039. The reliability of the scale is 0.109. The reliability of scale unbiased is 0.130. Guttman Split Half Coefficient is .744. Goodness of Fit Chi Square value international test is 111.149 degrees of freedom 4 Sig. 000. Log of determinant of constrained matrix is 1.405 and the constrained matrix is 2.721. The SD >1 is 4.89 and indicates that the scores are widely different, and scores are not clustered around the mean.

Reliability Indices suggest that the test items have good reliability. Cronbach’s Alpha is the mean of all (Flanagan–Rulon) split-half reliabilities. The result is exact if the test is split into two halves that are equal in size. This requires that the number of items is even, since odd numbers cannot be split into two groups of equal size. Alpha is approximately identical to the mean of all split-half reliabilities (Warrens, 2015)The correlation between forms national test is 0.306 indicating the test items are large and the randomly divided parts of the test are approximately equal. It assesses the performance of questions on the tests; therefore, the questions are good in discriminating bad students from good ones. Values should be between 0.2 and 0.39 to indicate good discrimination. (Exchange.com, 2020) Parallel Tests form reliability is administration of two similar homogeneous tests that have similar content, mental processes, length of test and difficulty level without being similar. These tests have nearly similar means and variances such as is the case with national and international tests. The obtained scores are correlated to give estimate of reliability called the coefficient of equivalence. The alternative forms tests determine stability of performance and equivalence of content. There is a negative covariance between some items which are not discriminating between the high ability and the low ability learners who were guessing correctly. Inter-item correlations examine the extent to which scores on one item are related to scores on all other items in a scale. It provides an assessment of item redundancy: the extent to which items on a scale are assessing the same content (Cohen, 1996) Ideally, the average inter-item correlation for a set of items should be between .20 and .40, suggesting that while the items are reasonably homogenous, they do contain sufficiently unique variance so as to not be isomorphic with each other (Piedmont, 2014)

**Table 3 *Parallel Test Reliability***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| National Test | IQ scores | Percentage | N of Students 85 | Mean Item difficulty |
| High Ability | 28.84 | 6.9% | 6 | 0.430 |
| Average Ability | 18.60 | 67.8% | 46 |  |
| Low Ability | 22.73 | 37.9% | 33 |  |
| International Test | IQ scores | Percentage | N of Students 87 | 0.264 |
| High Ability | 30.654 (3.4%) | 5.7% | 5 |  |
| Average Ability | 25.128 (1.1%) | 20.6% | 20 |  |
| Low Ability | 14.238 (25.3%) | 67.8% | 62 |  |
| Secondary School Exam | IQ scores | Percentage | N of Students 88 |  |
| High ability |  | 25.3% | 22 |  |
| Average Ability | 571.12(14.9%) | 71.2% | 63 |  |
| Low Ability | 29.80. | 3.4% | 3 |  |
| English Secondary school | IQ scores | Percentage | N of Students 91 |  |
| High Ability |  | 5.9% | 5 |  |
| Average Ability | 83.86 (40.2%) | 67.8% | 62 |  |
| Low ability | 9.90 | 6.9% | 6 |  |
|  | 46.88 | 18.4% | 18 |  |

**Table 4.** *Student grouping by performance -high ability, average ability, and low ability students*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | t-Test | Sig (Two tailed) | Mean Difference | The 95% confidence Interval of difference (Lower) | The 95% confidence Interval of difference (Upper) |
| National Test | 33.884 df (86) | .000 | 18.60920 | 17.5174 | 19.7010 |
|  |  |  |  |  |  |
| International Test | 33.597 df (86) | .000 | 19.71264 | 18.5462 | 20.8791 |
| Secondary School Exam | 63.573 df (83) | .000 | 778.40476 | 754.0513 | 802.7582 |
| English in Secondary School | 26.914 df (79) | .000 | 90.15000 | 83.4829 | 96.8171 |
| Table 5. *Paired Sample T-Test values* | | | | | |
|  | **Paired Sample Correlation** | **Sig (Two tailed)** | **Standard Error Mean** | **Mean** | **Standard Deviation** |
| National Test 1 International Test 1 | 1  0.011 | 0.406 | 0.32983  0.27289 | 12.6302  11.3385 | 7.23644  6.00358 |
| National Test 2 International Test 2 | -0.067  -0.077 | 0.061  0.022 | 0.27299  0.20779 | 9.0302  18.0043 | 5.88030  4.47600 |

**Table 6. *Test Data***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| NATIONAL TEST | | % Of correct answers | Most chosen Distractor | Item difficulty | Difficulty value of Test 19.91 High Difficulty <30% | Average of Items Left (no response=sum 25)  L (no response) |
| Qs 1. | | b 23.5% | a 76.5% | 0.23 (Difficult item-to be revised) | M .26 SD 0.442Inter-Item Correlation 1.000 |  |
| Qs 2 | | a 68.2% | b 29.4% | 0.68(Moderate difficulty very good item) | M .78 SD 0.414  Inter-Item Correlation  -0.002 | 2.4% |
| Qs 3 | | b 64.7% | a 32.9% | 0.64 (Moderate difficulty very good item) | M .74 SD 0.442  Inter-Item Correlation  -0.58 | 2.4% |
| Qs 4 | | b 5.9% | a 91.8% | 0.05 (Very difficult- to be discarded) | M .13 SD 0.333  Inter-Item Correlation 0.166 | 2.4% |
| Qs 5 | | b 35.3% | a 61.2% | 0.35 (Moderate difficulty very good item ) | M .41 SD 0.494  Inter-Item Correlation -0.179 | 2.4% |
| Qs 6 | | a 31.8% | e 34.1% | 0.31(Moderate difficulty very good item ) | M .36 SD 0.484  Inter-Item Correlation 0.34 | 2.4% |
| Qs 7 | | e 40% | b 38.8% | 0.4 (Very difficult-to be discarded) | M .45 SD 0.501  Inter-Item Correlation 0.236 |  |
| Qs 8 | | b 47.1% | c 21.2% | 0.47 (Moderate difficulty very good item) | M .55 SD 0.024  Inter-Item Correlation 0.024 | 1.2% |
| Qs 9 | | c 60% | b 22.4% | 0.6 (Very difficult-to be discarded) | M .68 SD 0.468 Inter-Item Correlation 0.073 | 1.2% |
| Qs 10 | | b 20.0% | e 35.3% | 0.2 (Very difficult-to be discarded) | M .51 SD 0.503  Inter-Item Correlation 0.64 | 1.2% |
| Qs 11 | | d 45.9% | a 16.5% | 0.45 (Moderate difficulty very good item) | M .80 SD 0.406 Inter-Item Correlation 0.45 |  |
| Qs 12 | | e 70.6% | a 15.3% | 0.70 (Moderate difficulty very good item) | M .64 SD 0.484 Inter-Item Correlation 0.073 |  |
| Qs 13 | | a 55.3% | e 21.2% | 0.55(Moderate difficulty very good item) | M .66 SD 0.477 Inter-Item Correlation 0.100 | 1.2% |
| Qs 14 | | d 58.8% | e 15.3% | 0.58(Moderate difficulty very good item) | M .49 SD 0.503 Inter-Item Correlation -0.12 |  |
| Qs 15 | | e 43.5% | d 30.6% | 0.43(Moderate difficulty very good item) | M .28 SD 0.454 Inter-Item Correlation 0.141 |  |
| Qs 16 | | c 24.7% | a 23% | 0.24 (Difficult item-to be revised) | M .26 SD 0.442 Inter-Item Correlation  -0.001 | 3.5% |
| Qs 17 | | b 23.5% | c 50.6% | 0.23 (Difficult item-to be revised) | M .19 SD 0.397 Inter-Item Correlation 0.36 | 1.2% |
| Qs 18 | | b 8.2% | c 37.6% | 0.08 (Very difficult-to be discarded) | M .20 SD 0.406 Inter-Item Correlation 0.019 | 1.2% |
| Qs 19 | | a 17.6% | e 34.1% | 0.17 (Very difficult- item to be discarded) | M .60 SD 0.492 Inter-Item Correlation 0.060 | 1.1% |
| Qs 20 | | e 51.8% | c 22.4% | 0.51(Moderate difficulty very good item) | M .19 SD 0.397 Inter-Item Correlation 0.092 | 2.4% |
| Qs 21 | | d 16.5% | e 36.5% | 0.16 (Very difficult item to be discarded) | M .24 SD 0.429 Inter-Item Correlation 0.003 | 2.4% |
| Qs 22 | | a 21.2% | c 40% | 0.21 (Difficult item-to be revised) | M .69 SD 0.464 Inter-Item Correlation 0.003 | 1.2% |
| Qs 23 | | c 56.5% | a 15.3% | 0.56 (Moderate difficulty very good item) | M .79 SD 0.464 Inter-Item Correlation 0.151 | 2.4% |
| Qs 24 | | a 68.2% | b & c 9.4% | 0.68(Moderate difficulty very good item) | M .79 SD 0.429 Inter-Item Correlation  -0.168 | 3.5% |
| Qs 25 | | d 42.4% | e 18.8% | 0.18 (Difficult item-to be revised) | M .49 SD 0.503 Inter-Item Correlation 0.60 | 3.5% |
| Qs 26  N=88  SEM= .586 | | d56.8% | a 26.1% | 0.56 (Moderate difficulty very good item) | M .69 SD0 .464 Inter-Item Correlation 1.000 |  |
| Qs 27 | | a 19.3% | b 65.9% | 0.19 Difficult item – to be revised) | M .24 SD 0.429 Inter-Item Correlation -0.90 | 1.1% |
| Qs 28 | | d 39.8% | c 38.6% | 0.39 (Moderate difficulty very good item) | M .49 SD 0.503 Inter-Item Correlation 0.010 |  |
| Qs 29 | | c 36.4 | a 31.8% | 0.36(Moderate difficulty very good item) | M .45 SD 0.501 Inter-Item Correlation 0.013 | 1.1% |
| Qs 30 | | b 34.1% | e 23.9% | 0.34(Moderate difficulty very good item) | M .42 SD 0.496 Inter-Item Correlation -0.032 | 1.1% |
| Qs 31 | | b 71.6% | c 9.1% e 9.1% | 0.71 (Easy item-to be revised) | M .85 SD 0.357 Inter-Item Correlation 0.626 | 1.1% |
| Qs 32 | | a 55.7% | d 23.9% | 0.55 (Moderate difficulty very good item) | M .70 SD 0.459 Inter-Item Correlation 0.973 |  |
| Qs 33 | | c 28.4% | b 37.5% | 0.28 (Difficult item-to be revised) | M .35 SD 0.480 Inter-Item Correlation  -0.025 |  |
| Qs 34 | | e 9.1% | a 35.2% | 0.09 (Very difficult item-to be discarded) | M .17 SD 0.378 Inter-Item Correlation 0.039 |  |
| Qs 35 | | d 19.3% | e 23.9% | 0.19 (Very difficult item –to be discarded) | M .24 SD 0.429 Inter-Item Correlation  -0.90 |  |
| Qs 36 | | a 15.9% | c 29.5% | 0.15 (Difficult item –to be revised | M .73 SD 0.448 Inter-Item Correlation  -0.090 |  |
| Qs 37 | | b 36.4% | c 27.3% | 0.36 (Moderate difficulty very good item) | M .13 SD 0.333 Inter-Item Correlation 0.251 |  |
| Qs 38 | | a 15.9% | b 35.2% | 0.15 (Very difficult item-to be discarded) | M .75 SD 0.435  Inter-Item Correlation 0.914 |  |
| Qs 39 | | c 18.2% | d 53.4% | 0.18 (Difficult item to be revised) | Inter-Item Correlation -0.76 | 1.1% |
| Qs 40 | | a 61.4% | c 15.9% | 0.61 (Moderate difficulty very good item) | Inter-Item Correlation 0.466 | 1.1% |
| Qs 41 | | a 22.7% | c 37.5% | 0.22 (Difficult item-to be revised) | Inter-Item Correlation 0.810 |  |
| Qs 42 | | b 28.4% | a 34.1% | 0.28 (Difficult item-to be revised) | Inter-Item Correlation 0.759 |  |
| Qs 43 | | a 18.2% | c 22.7% | 0.18 (Difficult item-to be revised) | Inter-Item Correlation -0.466 |  |
| Qs 44 | | c 6.8% | d 31.8% | 0.06 (Very difficult item-to be discarded) | Inter-Item Correlation  0.421 | 1.1% |
| Qs 45 | | a19.3% | d 43.2% | 0.19 (Difficult item-to be revised) | Inter-Item Correlation 0.035 | 2.3% |
| Qs 46 Dialogue Writing | | 2 /5 29.5%  3/5 20.5%  4/5 19.3%  5/5 12.5% |  | 0.875 (Easy item- to be revised) | Inter-Item Correlation  -1.06 | 12.5% |
|  | **INTERNATIONAL TEST DATA** | | | | | |
|  | | % of correct answers |  | Item Difficulty |  | Question was left by students (no response) |
| Qs 47 | | 56.8% | Supply type Prefix: fore | 0.56 (Moderate difficulty very good item) |  | 43.2% |
| Qs 48 | | 53.4% | In | 0.46 (Moderate difficulty very good item) |  | 46.6% |
| Qs 49 | | 54.5% | Dem | 0.45(Moderate difficulty very good item) |  | 45.5% |
| Qs 50 | | 55.7% | Inter | 0.55(Moderate difficulty very good item) |  | 42% |
| Qs 51 | | 14.8% | Mal | 0.14 (Very difficult item-to be discarded) | M 56 SD 0.499  Inter-Item Correlation 1.000 | 79.5% |
|  | |  | **Most chosen Distractor** |  |  |  |
| Qs 52 | | d 22.7% | e 28.4% | 0.22 (Difficult item-to be revised) | M .26SD 0.444 Inter-Item Correlation 0.055 | 4.5% |
| Qs 53 | | d 37.42% | a 22.7% | 0.42 (Moderate difficulty very good item) | M .46 SD 0.567 Inter-Item Correlation  -0.063 |  |
| Qs 54 | | a 10.2. % | c 36.4% | 0.10 (Very difficult item-to be discarded) | M .37 SD 0.485  Inter-Item Correlation 0.239 |  |
| Qs 55 | | e 12.5 % | c 28.4% | 0.12 (Very difficult item-to be discarded) | M .21 SD 0.631  Inter-Item Correlation. 0.180 |  |
| Qs 56 | | d 13.6 % | c 43.2% | 0.13 (Very difficult item-to be discarded) | M .15 SD 0.359 Inter-Item Correlation 0.044 |  |
| Qs 57 | | c 51.1% | b 18.2% | 0.51(Moderate difficulty very good item) | M .59 SD 0.495 Inter-Item Correlation  -0.34 |  |
| Qs 58 | | b 39.8% | d 22.7% | 0.39(Moderate difficulty very good item) | M .47 SD 0.502  Inter-Item Correlation  -0.051 | 1.1% |
| Qs 59 | | b 21.1% | d 28.4% | 0.21(Difficult item-to be revised) | M .24 SD 0.430  Inter-Item Correlation -0.99 | 1.1% |
| Qs 60 | | a 43.2% | b 31.8% | 0.43 (Moderate difficulty very good item) | M .47 SD 0.502  Inter-Item Correlation  -0.97 | 1.1% |
| Qs 61 | | b 56.8% | a 14.8% | 0.56(Moderate difficulty very good item) | M .67 SD 0.474  Inter-Item Correlation -0.33 | 1.1% |
| Qs 62 | | e 12.5% | b 45.5% | 0.125 (Very difficult item-to be discarded) | M.15 SD 0.359 Inter-Item Correlation -0.86 |  |
| Qs 63 | | e 20.5% | b 34.1% | 0.20 (Difficult item-to be revised) | M .24 SD 0.430 Inter-Item Correlation 0.009 |  |
| Qs 64 | | e 13.6% | b 30.7% | 0.13 (Very difficult item-to be discarded) | M .16 SD 0.370  Inter-Item Correlation  -0.56 | 1.1% |
| Qs 65 | | b 27.3% | c 29.5% | 0.27 (Difficult item – to be revised) | M.25 SD 0.437 Inter-Item Correlation  -0.21 | 1.1% |
| Qs 66 | | b 30.7% | a 26.1% | 0.30 (Moderate difficulty very good item ) | M .37 SD 0.485  Inter-Item Correlation  -0.001. | 1.1% |
| Qs 67 | | b 18.2% | c 37.42% | 0.18 (Difficult item-to be revised) | M .24 SD 0.430  Inter-Item Correlation 0.009 | 1.1% |
| Qs 68 | | c 23.9% | b 34.1% | 0.23 (Difficult item to be revised) | M.29 SD 0.455  Inter-Item Correlation  -0.55 | 1.1% |
| Qs 69 | | e 19.3% | c 44.3% | 0.19 (Difficult item –to be revised) | M .24 SD 0.430  Inter-Item Correlation .0.45 | 1.1% |
| Qs 70 | | e 5.7% | c 34.1% | 0.05 (Very difficult item – to be discarded) | M .07 SD 0.255  Inter-Item Correlation  -0.126 | 1.1% |
| Qs 71 | | d 10.2% | c 40.9% | 0.10 (Very difficult item-to be discarded) | M .13 SD 0.334  Inter-Item Correlation  -0.14 |  |
| Qs 72  N=91 | | e 37.4% | b 26.4% | 0.37 (Moderate difficulty very good item) | M .44 SD 0.499  Inter-Item Correlation  -0.28 | 3.3% |
| Qs 73 | | d 51.6% | e 17.6% | 0.51(Moderate difficulty very good item) | M.71 SD 0.455  Inter-Item Correlation 0.055 |  |
| Qs 74 | | a 35.2% | e 18.7% | 0.36(Moderate difficulty very good item) | M.40 SD 0.493  Inter-Item Correlation 0.014 |  |
| Qs 75 | | c 46.2% | a 22% | 0.46(Moderate difficulty very good item) | M.54 SD 0.501  Inter-Item Correlation  -0.68 |  |
| Qs 76 | | b 41.8% | a 15.4%  c 15.4% | 0.41(Moderate difficulty very good item) | M .45 SD 0.500 Inter-Item Correlation  -1.00 |  |
| Qs 77 | | c 6.6% | a 17.6% | 0.06 (Very difficult item-to be discarded) | M .08 SD 0.268 Inter-Item Correlation  -0.013 |  |
| Qs 78 | | d 13.2% | b 25.3% | 0.13 (Very difficult item- to be discarded) | M .15 SD 0.363 Inter-Item Correlation  -0.374 |  |
| Qs 79 | | c 56% | b 18.7% | 0.56 (Moderate difficulty very good item) | M .62 SD 0.489  Inter-Item Correlation 0.281 |  |
| Qs 80 | | b 59.3% | c 13.2%  e 13.2% | 0.05 (Very difficult item- to be discarded) | M 60 SD 0.492 Inter-Item Correlation 0.520 |  |
| Qs 81 | | d 22% | e 29.7% | 0.21 (Difficult item- to be revised) | M .24 SD 0.431 Inter-Item Correlation 0.144 |  |
| Qs 82 | | a 28.6% | c 20.9% | 0.28 (Difficult item –to be revised) | M .35 SD 0.480 Inter-Item Correlation  -0.125 |  |
| Qs 83 | | d 5.5% | a 47.3% | 0.05 (Very difficult item-to be discarded) | M 0.10 SD 0.300 Inter-Item Correlation  -0.120 |  |
| Qs 84 | | d 28.6% | c 27.5% | 0.28 (Difficult item-to be revised) | M .31 SD 0.464 Inter-Item Correlation 0.76 |  |
| Qs 85 | | b 20.9% | d 34.1% | 0.20 (Difficult item-to be revised) | M .25 SD 0.437 Inter-Item Correlation  -0.106 |  |
| Qs 86 | | d 16.5% | b 46.2% | 0.16 (Very difficult item- to be discarded) | M .19 SD 0.392 Inter-Item Correlation 0.361 |  |
| Qs 87 | | a 82.4% | b 15.4% | 0.82 (Easy item-to be revised) | M .77 SD 0.422 Inter-Item Correlation  0 .214 |  |
| Qs 88 | | a 45.1% | b 53.8% | 0.45 (Moderate difficulty very good item) | M .64 SD 0.483 Inter-Item Correlation 0.436 |  |
| Qs 89 | | a 70.3% | b 28.6% | 0.70(Moderate difficulty very good item) | M .64 SD 0.483 Inter-Item Correlation 0.577 |  |
| Qs 90 | | a 58.2% | 40.7% | 0.58(Moderate difficulty very good item) | M .78 SD 0.416 Inter-Item Correlation 0.733 |  |
| Qs 91 | | a 51.6% | b 46.2% | 0.51(Moderate difficulty very good item) | M .65 SD 0.480 Inter-Item Correlation  -0.366 |  |
| Qs 92 | | a 54.9% | b 41.8% | 0.54(Moderate difficulty very good item) | M .60 SD 0.492 Inter-Item Correlation  -0.319 |  |
| Qs 93 | | a 64.8% | b 33% | 0.64(Moderate difficulty very good item) | M .60 SD 0.492 Inter-Item Correlation 0.376 |  |
| Qs 94 | | a 72.5% | b 25.3% | 0.72 (Easy item-to be revised) | M .76 SD 0.431 |  |
| Qs 95 | | a 53.8% | b 44% | 0.53 (Moderate difficulty very good item) | M .78 SD 0.416 |  |
| Qs 96 | | a 57.1% | b 28.6% | 0.57(Moderate difficulty very good item) | M .55 SD 0.500 |  |
| Item Bank  Indices of difficulty (Hopkins and Antes) | | **Very Easy items**  **0.71-0.85** | **Moderate Difficulty Very Good items**  **0.30-0.70** | **Difficult items to be revised**  **0.15-0.29** |  | **Very difficult items to be discarded**  **0.14 – below** |
|  | | **3** | **43** | **23** |  | **24** |

Heteroscedasticity in scores of Secondary school exam and the English exam of secondary school implies students’ performance is greatly varied and divergent. Heteroskedasticity (or heteroscedasticity) happens when the standard deviations of a predicted variable, monitored over different values of an independent variable or as related to prior time periods, are non-constant (Hayes, 2020)

**Table 7: *Secondary School Exam and Marks in Subject of English***

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Pearson Correlation values  sig (2 tailed) | Marks in National Tests n=87 | Marks in International Tests n=87 | Marks in Secondary School Exam n=84 | Marks in English Secondary school  n=80 |
| National Test | 1 | .497\* | .223 | -0.11 |
|  |  | sig 000 | sig -042 | sig -923 |
| International Test | .497\* | 1 | .045 | -154 |
|  | sig 000 |  | .684 | .173 |
| Secondary School Exam | .223\*  sig .042 | .045  sig .684 | 1 | 313\*\*  sig .005 |

**2 Phase 2 Assessment tool (2) Teacher evaluation by students**. Learners’ satisfaction with blended learning was recorded on a Likert scale questionnaire scale ranging from 5 strongly agree, 4 agree, 3neutral, 2 disagree and 1 strongly disagree.

**Table 8**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Item-Total Statistics for 3 Variables** | | | | |
|  | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Cronbach's Alpha if Item Deleted |
| **Directive Teacher Method** | 8.03 | 1.398 | 0.575 | 0.758 |
| **Communicative Teacher Method** | 7.73 | 1.578 | 0.656 | 0.670 |
| **Students’ Development** | 7.83 | 1.451 | 0.634 | 0.683 |

The Correlation coefficient Spearman’s Rho for the nonparametric ordinal Likert scale questionnaire data is 1.000 sig.2-tailed .000 for the Directive Teacher method but 0.586 for the Communicative teacher method. The Communicative Teacher method average value 4.0100 (SD 0.627, Std. Error Mean 0.06276) is higher than the directive teacher average of 3.7900(SD 0.74597, Std. Error Mean 0.0746) Valid no. of responses is 100 (75.2%), the excluded is 33 (24.8%) Student responses indicate the teachers’ methodology is communicative Questionnaire Cronbach Alpha reliability is *0.741*. This instrument reliability is good though, subscale adequate level of inter-item reliability is 0.598 directive and 0.598 communicative and can be improved. There is negative covariance between these two variables which indicates that these two methodologies move in opposite directions. The Npar test results indicate the median value for the Directive Teacher Method to be 4.00 (n=100, Chi-square 10.092 df 3Asymp. sig .018). The valid response is 75.2%. Since the p value >0.05 the null hypothesis of equal population medians is retained. The Communicative Teacher method median value is 4.00. (n=108, Chi-square 29.596 Asymptomatic sig.000). The response is 81.2%. Since the p value is < 0.05 the null hypothesis of equal population means is rejected. The Cronbach’s Alpha reliability of the three new variables is *.780* which is a good value for alpha.

**Conclusion**

This quantitative research compared two methods of learning - the conventional secondary school exam and English subject scores with the national and international tests in the blended learning environment. Teacher’s evaluation by students highlighted teachers’ skills and competencies. The principles of measurement provide a framework in which tests from outside the school maybe adopted (published tests), tests within the school may also be developed and test scores can be meaningfully interpreted and reported to provide meaningful and dependable results. This practice should be continued for its numerous benefits.

**References**

Arter, J. A. (2010). *Interim benchmark assessments: Are we getting our eggs in the right basket?* Paper presented at the Annual meeting of the National Council of Measurement in Education., Denver, CO.

Basilaia, G., & Kvavadze, D. (2020). *Transition to online education in schools during a SARS-CoV-2 coronavirus (Covid-19) pandemic in Georgia.* Pedagogical Research, 5(4), 1-9.

Boyle, J. A. a. C. (2020). *Including into What? Reigniting the ‘Good Education’ Debate in an Age of Diversity.* In Inclusive Education: Global Issues and Controversies (pp. 15-34): <https://brill.com/view/book/edcoll/9789004431171/BP000014.xml>.

Cook, V. (2016). *Second Language Learning and Language Teaching.* In Language Arts & Disciplines USA: Routledge.

Doughlas, M. (1960). The Human Side of Enterprize. New York: McGraw -Hill.

Douglas Fisher, N. F. a. C. R. (2008). *Content-Area Conversations: How to Plan Discussion-Based Lessons for Diverse Language Learners*.

Garcia. (2002). Student cultural diversity; understanding the meaning and meeting the challenge. Boston: MA: Houghton Mifflin.

Hayes,A.(2020).Heteroskedasticity.Retrievedfrom <https://www.investopedia.com/terms/h/heteroskedasticity.asp>

Khan., P. Z.-u.-R. (2015). In Intermediate English Simple Grammar and Composition (Part II) Urdu Bazar, Lahore: Pakistan: Simple Publications.

Kosslyn, S. (2015). *Cognitive Psychology Mind and Brain*. Pearson India Education Services.

Hayes, D, Martin M., Lingard, B (2006). *Teachers and Schooling Making A Difference: Productive pedagogies Assessment and Performance Studies in education.* In D. N. A. Hayes (Ed.), Education: Allen & Unwin.

Mailizar, A., A., Maulina, S., & Bruce, S. (2020). *Secondary school mathematics teachers’ views on e-learning implementation barriers during the Covid-19 pandemic: The case of Indonesia*. Eurasia Journal of Mathematics, Science and Technology Education, 16(7) (em1860).

Miller, D. (1990). Organizational configurations: Cohesion, change, prediction Human Relations, 43(8), 771-789.

Shrum, G. (2015). Teacher's Handbook, Contextualized Language Instruction.

Smith, R. M. (1982*). Husserl and Intentionality: A Study of Mind, Meaning, and Language*. In Springer Science & Business Media. https://books.google.com.sg/books?id=SaCQHaPPz7UC.

Snow, F. (2000). What teachers need to know about language. Special Report. Tavassoli\*, N. K. K. (2020). The Comparative Effect of Dynamic vs. Diagnostic Assessment on EFL Learners’ Speaking ability Research in English Language Pedagogy, 8(2): 223-241 doi:10.30486/RELP.2019.1878561.1155

TEST, N. D. R. (2020). NELSON-DENNY READING TEST (NDRT) BREAKDOWN. ONLINE PRACTICE TESTS COLLEGE ADMISSIONS AND PLACEMENT.

Shrestha. P (2017) “Ebbinghaus Forgetting Curve," in Psychestudy <https://www.psychestudy.com/cognitive/memory/ebbinghaus-forgetting-curve>.

Wolf, D., Bixby, J., Glenn III, J. and Gardner, H. (1991). To use their minds well: new forms of student assessment. Review of Research in Education, 17, 31-74