**Manchester Exam**

**MANCHESTER EXAM - Approval for use in the United States**

Prepared by

**Eliza P. Aimone, B.S. Data Analytics**

**Executive Summary**

This white paper rigorously evaluates the Manchester Exam as a viable alternative to established English as a Foreign Language (EFL) assessments such as TOEFL and IELTS within the American education system. The objective is to substantiate the Manchester Exam's efficacy in accurately measuring English language proficiency among non-native speakers, addressing the evolving needs of both students and educational institutions. This document demonstrates the Manchester Exam's potential as a robust tool for assessing EFL competencies in the United States by critically analyzing its testing methodology, alignment with contemporary linguistic pedagogies, and comparability with existing standards.

The Manchester Exam is a comprehensive English proficiency test designed to evaluate the English skills of non-native speakers for admission to academic institutions globally. It encompasses six sections—Grammar, Comprehension, Reading, Listening, Writing, and Speaking—each targeting specific language skills essential for academic and professional success. Key features of the Manchester Exam include its flexibility and accessibility, allowing it to be taken online from any location with a stable internet connection. It employs robust security measures such as live proctoring and AI monitoring to ensure test integrity. Additionally, it offers a rapid scoring service for U.S. institutions, with results delivered within 24 hours, enhancing admission process efficiency. Internationally recognized and validated by reputable organizations like Tennessee Tech University and TLM, the Manchester Exam is developed through extensive research and validation by American and British academics, ensuring its credibility and alignment with current educational standards.

To further enhance its utility and recognition, INTO USF proposes the inclusion of the Manchester Exam as an accepted measure of English proficiency for international graduate students. Specifically, applicants may demonstrate proficiency by providing an overall score of 301 or higher on the Manchester Exam, with minimum sub-scores of 75 in each section. This addition to the university’s graduate handbook will offer greater flexibility and access for international students, aligning to accommodate diverse linguistic backgrounds and supporting a more inclusive admissions process.

Applicants from countries where English is not the official language must also demonstrate proficiency in English in one of the following ways: a. By providing scores of 79 or higher on the internet-based Test of English as a Foreign Language (TOEFL iBT) b. By providing a score of 6.5 or higher on the International English Language Testing System (IELTS) c. By providing a score of 53 or higher on the Pearson Test of English Academic (PTE-A) d. By providing an overall score of 176 or higher (with minimum sub-scores of 169) on the INTO English Language Assessment (IELA) e. By providing an overall score of 176 or higher (with minimum sub scores of 169) on the Cambridge English First (FCE) f. By providing an overall score of 301 or higher (with minimum sub scores of 75) on the Manchester Exam g. By earning a 153 (or equivalent) score on the GRE Verbal exam h. By earning a baccalaureate or higher degree at a regionally accredited institution in the U.S. i. Applicants who earn a baccalaureate or equivalent degree at a foreign institution where English is the language of instruction (for the institution and not just the major) may meet this requirement. However, other factors (including test scores) will also be considered. Medium of Instruction must be documented on the transcript or an official Certificate of Medium of Instruction from the Institution.

The analysis conducted in this study highlights the Manchester Exam's strong potential to serve as an effective and comprehensive tool for assessing English language proficiency. Its innovative features, combined with a focus on both academic and practical language skills, make it a valuable addition to the landscape of EFL assessments. The flexibility, accessibility, and rapid scoring of the Manchester Exam enhance its appeal to both students and educational institutions, promising to streamline the admissions process while maintaining high standards of test integrity and fairness.

This white paper concludes that the Manchester Exam can address the diverse needs of international students and educational institutions, providing a robust and reliable measure of English language proficiency. As the Manchester Exam continues to gain recognition and undergo further validation, it is poised to become a trusted and widely adopted assessment tool in the United States and beyond

**Abbreviations**

|  |  |
| --- | --- |
| Abbreviation | Full Form |
| MEXAM | Manchester Exam |
| TOEFL | Test of English as a Foreign Language |
| IELTS | International English Language Testing System |
| PTE-A | Pearson Test of English Academic |
| IELA | INTO English Language Assessment |
| FCE | First Certificate in English (Cambridge English First) |
| EFL | English as a Foreign Language |
| AI | Artificial Intelligence |
| USF | University of South Florida |
| ETS | Educational Testing Service |
| TLM | The Learning Machine |
| TESOL | Teaching English to Speakers of Other Languages |
| CEFR | Common European Framework of Reference for Languages |
| ANOVA | Analysis of Variance |
| df | Degrees of Freedom |
| Sig. | Significance |
| R² | Coefficient of Determination |
| Std. Error | Standard Error |
| N | Number of Observations |
| GPA | Grade Point Average |
| APA | American Psychological Association |
| L2 | Second Language |
| ITA | International Teaching Assistant |
| MOI | Medium of Instruction |

**Introduction and Background**

The Manchester Exam is a comprehensive English proficiency test designed to evaluate the English skills of non-native speakers for admission to academic institutions worldwide. It encompasses six sections: Grammar, Comprehension, Reading, Listening, Writing, and Speaking, each targeting specific language skills. The exam's structure thoroughly assesses an individual's ability to understand and use English in various contexts, making it highly relevant for academic and professional settings. The Grammar section involves fill-in-the-gap exercises and identifying incorrect words, while the Comprehension section includes tasks like naming pictures and filling in missing letters. The Reading section requires participants to read passages and answer questions to demonstrate their comprehension, and the Listening section involves responding to audio clips. The Writing section includes describing pictures and writing essays, and the Speaking section involves reading paragraphs aloud and answering spoken questions (Bachman & Palmer, 1996; Xi, 2008).

The purpose of this study is to evaluate the Manchester Exam as a viable alternative to established English as a Foreign Language (EFL) assessments such as TOEFL and IELTS within the American education system. With the evolving needs of students and educational institutions, it is essential to explore new tools that can accurately measure English language proficiency among non-native speakers.

The Manchester Exam is a comprehensive English proficiency test designed to evaluate various language skills necessary for academic and professional success. It includes six sections: Grammar, Comprehension, Reading, Listening, Writing, and Speaking. Each section targets specific language skills, making the exam a thorough assessment of an individual's ability to understand and use English in different contexts. The exam is flexible and accessible, allowing it to be taken online from any location with a stable internet connection. It employs robust security measures such as live proctoring and AI monitoring to ensure the integrity of the test. Additionally, it offers a rapid scoring service for U.S. institutions, with results delivered within 24 hours, enhancing the efficiency of the admission process.

The significance of evaluating the Manchester Exam lies in its potential to address the evolving needs of both students and educational institutions. As the demand for high-quality and accessible English language assessments grows, it is crucial to identify tools that can provide accurate and reliable measures of English proficiency. The Manchester Exam's design and methodology aim to align with contemporary linguistic pedagogies and educational standards, making it a promising candidate for widespread adoption.

The objectives of this study are to critically analyze the Manchester Exam's testing methodology, assess its alignment with current linguistic theories and educational practices, and compare its effectiveness with existing standards such as TOEFL and IELTS. By doing so, we aim to substantiate the Manchester Exam's efficacy in measuring English language proficiency and its potential to serve as a robust tool for assessing EFL competencies in the United States. This evaluation will provide insights into the exam's strengths and areas for improvement, guiding educational institutions in making informed decisions about adopting the Manchester Exam as part of their assessment processes.

One of the primary advantages of the Manchester Exam is its flexibility and accessibility. The exam can be taken online from any location with a stable internet connection, providing significant convenience for test-takers, especially those in remote areas. Additionally, the exam employs robust security measures, including live proctoring and AI monitoring, to ensure the integrity of the test. For U.S. institutions, it offers a rapid scoring service, with results delivered within 24 hours, enhancing the efficiency of the admission process. The exam is internationally recognized and validated by reputable organizations such as Tennessee Tech University and TLM, ensuring its credibility and alignment with current educational standards. The development of the Manchester Exam involved extensive research and validation by American and British academics with advanced degrees in TESOL and linguistics.

1. **Literature Review**

The Test of English as a Foreign Language (TOEFL) is a prominent English language proficiency test developed by the Educational Testing Service (ETS). It evaluates four essential language skills: Reading, Listening, Speaking, and Writing. Each section is designed to assess different aspects of English proficiency, providing a comprehensive measure of a candidate’s academic ability to understand and use English effectively.

The TOEFL iBT (Internet-Based Test) employs a combination of automated and human evaluations, particularly in the Speaking and Writing sections, to ensure a fair and consistent measure of language proficiency. Each section is scored on a scale of 0 to 30, with a total score range of 0 to 120. Research indicates that TOEFL scores are reliable indicators of academic success in English-speaking institutions. However, the test has faced criticism for its emphasis on academic English, which may not fully reflect a test-taker's ability to use English in everyday contexts. Additionally, the test's cost and accessibility can be prohibitive for some test-takers, particularly those from lower socioeconomic backgrounds.

Overview of IELTS and Its Methodology

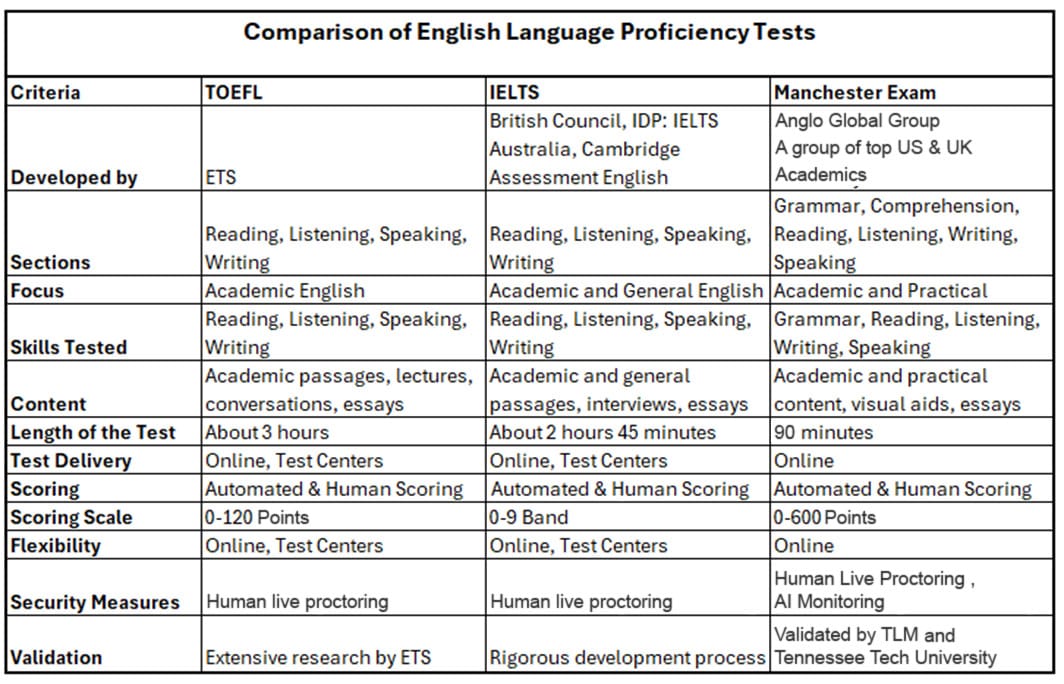
The International English Language Testing System (IELTS) is another widely recognized English language proficiency test, jointly managed by the British Council, IDP: IELTS Australia, and Cambridge Assessment English. IELTS assesses the English language skills of non-native speakers who intend to study or work in environments where English is the language of communication. The test comprises four sections: Listening, Reading, Writing, and Speaking, each designed to evaluate different aspects of language proficiency.

IELTS is known for its strong emphasis on real-world communication skills, ensuring that test-takers are prepared for practical English usage, not just academic tasks. The test is scored on a nine-band scale, with each band corresponding to a specific level of English proficiency. Continuous research and updates ensure that IELTS remains a robust tool for assessing English language proficiency, aligning with international standards. However, IELTS has faced criticisms regarding its scoring consistency and test administration. Variability in speaking test scores due to differences in examiner judgment can affect the overall reliability of the test. Additionally, the test's time constraints can add unnecessary pressure, potentially impacting performance, especially for non-native speakers who may require more time to process and respond in English.

Detailed Comparison with the Manchester Exam

The Manchester Exam is a relatively new entrant in the realm of English language proficiency assessments. It was developed to provide a comprehensive measure of English language proficiency, addressing both educational and practical language skills. The exam includes six sections: Grammar, Comprehension, Reading, Listening, Writing, and Speaking, each targeting specific linguistic competencies essential for success in both educational and real-world settings.

One of the distinctive features of the Manchester Exam is its integration of academic and practical content, ensuring that test-takers are equipped for both educational environments and practical everyday communication. The exam emphasizes accessibility and flexibility, offering the test exclusively online. This feature allows candidates to take the exam from any location with a reliable internet connection, making it particularly advantageous for test-takers in remote areas or those with limited access to traditional test centres. The exam employs robust security measures, including live proctoring and AI monitoring, to ensure the integrity and fairness of the testing process.

****

**TOEFL Test**

The Test of English as a Foreign Language (TOEFL) is a prominent English language proficiency test developed by the Educational Testing Service (ETS) and is primarily used for admission to English-speaking universities. The TOEFL iBT evaluates four essential language skills: Reading, Listening, Speaking, and Writing, each designed to assess different aspects of English proficiency. This comprehensive structure evaluates a candidate’s academic ability to understand and use English effectively. Research indicates that TOEFL scores are reliable indicators of academic success inEnglish-speaking institutions. For instance, a study by Inoue (2017) found that TOEFL iBT scores moderately correlated with local screening exams for international teaching assistants (ITAs), with listening scores being better predictors of teaching competence than speaking scores. Despite its strengths, TOEFL has faced criticism for its emphasis on academic English, which may not fully reflect a test-taker's ability to use English in everyday contexts. Additionally, the test's cost and accessibility can be prohibitive for some test-takers, particularly those from lower socioeconomic backgrounds (Educational Testing Service, 2021).

The TOEFL scoring algorithm combines automated and human evaluations, particularly in the speaking and writing sections, to ensure a fair and consistent measure of language proficiency. Each section of the TOEFL iBT is scored on a scale of 0 to 30, with a total score range of 0 to 120. This dual approach aims to balance the objectivity of machine scoring with the nuanced understanding of human raters (Educational Testing Service, 2021). Furthermore, continuous updates and extensive research underpinning the TOEFL test help maintain its relevance and reliability as a measure of English language proficiency.

**IELTS Test**

The International English Language Testing System (IELTS) is another widely recognized English language proficiency test, jointly managed by the British Council, IDP: IELTS Australia, and Cambridge Assessment English. IELTS assesses the English language skills of non-native speakers who intend to study or work in environments where English is the language of communication. The test comprises four sections: Listening, Reading, Writing, and Speaking, each designed to evaluate different aspects of language proficiency. The IELTS test is known for its strong emphasis on real-world communication skills, which helps ensure test-takers are prepared for practical English usage, not just academic tasks (Mohammad & Ghazali, n.d.). Research supports the validity and reliability of IELTS, noting that the rigorous development process ensures high assessment standards (Inoue, 2017).

However, IELTS has faced criticisms regarding its scoring consistency and test administration. A study by Fulcher (2003) pointed out potential variability in speaking test scores due to differences in examiner judgment, affecting the overall reliability of the test. Additionally, some researchers argue that the test's time constraints can add unnecessary pressure, potentially impacting performance, especially for non-native speakers who may require more time to process and respond in English (Cambridge English, 2004). IELTS scores are reported on a nine-band scale, with each band corresponding to a specific level of English proficiency, providing a detailed profile of a candidate's abilities across the four skills. Continuous research and updates ensure that IELTS remains a robust tool for assessing English language proficiency, aligning with international standards, and meeting the needs of various stakeholders, including educational institutions and employers.

**Manchester Exam**

The Manchester Exam is a relatively new entrant in the realm of English language proficiency assessments, designed to evaluate the English skills of non-native speakers for admission to academic institutions worldwide. TLM and Tennessee Tech University developed this exam to provide a comprehensive measure of English language proficiency that addresses educational and practical language skills. It includes six sections: Grammar, Comprehension, Reading, Listening, Writing, and Speaking, each targeting specific linguistic competencies essential for success in both educational and real-world settings (Manchester Exam, 2021). One of the distinctive features of the Manchester Exam is its integration of academic and practical content, which helps ensure test-takers are equipped for educational environments and practical, everyday communication.

The Manchester Exam emphasizes accessibility and flexibility, offering the test exclusively online. This feature allows candidates to take the exam from any location with a reliable internet connection, making it particularly advantageous for test-takers in remote areas or those with limited access to traditional test centres. Additionally, the exam employs robust security measures, including live proctoring and AI monitoring, to ensure the integrity and fairness of the testing process (Manchester Exam, 2021). Validation and credibility are key strengths of the Manchester Exam, validated by both US and UK academics, ensuring its alignment with high educational standards and reliability as an assessment tool. Extensive research and validation efforts have been conducted to ensure that the exam accurately measures the English language proficiency of non-native speakers (Educational Testing Service, 2021).

1. **Methodology**

The methodology for English as a Foreign Language (EFL) tests typically involves a structured approach to assessing multiple language skills, including reading, listening, speaking, and writing. Each section is designed to evaluate specific competencies and provide a comprehensive assessment of a test-taker’s proficiency in English. The design and administration of these tests are grounded in rigorous research and psychometric principles to ensure validity, reliability, and fairness (Weir, 2005).

**Scoring Methodologies**

Scoring methodologies for EFL tests vary significantly depending on the design and objectives. Commonly, tests like TOEFL and IELTS employ a combination of automated and human scoring. Automated scoring systems, particularly for reading and listening sections, use algorithms to evaluate responses against a predefined set of criteria, ensuring consistency and objectivity (Wang, 2010). Human raters are often employed for speaking and writing sections due to the complexity and nuance involved in these skills. Human raters are trained to assess various aspects of language use, including coherence, grammatical accuracy, vocabulary usage, and pronunciation. Using human raters allows for a more nuanced and context-sensitive evaluation, which automated systems might not fully capture (Xi, 2008).

The Manchester Exam emphasizes hand scoring for the speaking and writing sections, recognizing this method's advantages in terms of reliability and validity. Hand scoring allows raters to consider responses' context, nuance, and complexity, providing a more comprehensive assessment of test-taker's abilities. This approach is particularly beneficial for evaluating spoken language, where intonation, stress, and rhythm play crucial roles in communication (Knoch & Elder, 2010).

**Weighting of Questions**

The weighting of questions in EFL tests is carefully calibrated to reflect the relative importance of different language skills and ensure a balanced assessment. Typically, sections such as reading and listening may have more questions to broadly sample abilities, while speaking and writing sections. However, fewer in number are given significant weight due to their complexity and the depth of skills they assess (Kane et al., 1999). In the Manchester Exam, the weighting of questions is designed to balance the comprehensive assessment of academic and practical language skills. Each section contributes proportionately to the overall score, particularly emphasising productive skills like speaking and writing, which are critical for academic success. This balanced weighting ensures that test-takers are evaluated holistically, capturing a wide range of language abilities (Manchester Exam, 2021).

**Training for Scoring Personnel**

Training for scoring personnel is critical to ensuring the reliability and validity of hand-scored sections in EFL tests. Raters undergo extensive training to become familiar with scoring rubrics, practice with sample responses, and receive calibration to align their scoring with established standards. This process includes regular reliability checks and recalibration sessions to maintain consistency and objectivity over time (Lumley, 2002). Its rigorous training program for raters reflects the Manchester Exam’s commitment to high-quality scoring. Scoring personnel are selected based on their expertise in linguistics and language teaching, and they undergo continuous professional development to stay updated with best practices in language assessment. This ensures that the scores they assign are accurate and fair and reflect the test-taker's true abilities (Manchester Exam, 2021).

**Test Question Creation Methodologies**

Creating test questions for EFL assessments involves multiple stages of development, review, and validation to ensure that the test is fair and effective. The methodology for crafting these questions is grounded in rigorous research and psychometric principles, aiming to accurately measure non-native speakers' language proficiency. This process includes the generation of items, pilot testing, item analysis, and continuous refinement based on statistical data and expert feedback (Bachman & Palmer, 1996).

The initial stage of test question creation involves item generation, where experienced item writers develop questions based on specific test specifications and content frameworks. These specifications outline the skills to be assessed, the difficulty level, and the format of the questions. The generated items are then subjected to pilot testing administered to a representative sample of test-takers. This stage helps identify any potential issues with the questions, such as ambiguity, cultural bias, or inappropriate difficulty levels (Alderson, 1993). Pilot testing is followed by item analysis, which uses statistical techniques to evaluate the performance of each question. This analysis includes examining item difficulty, discrimination indices, and other psychometric properties to ensure that each question accurately measures the intended construct. Questions that do not meet the required standards are revised or discarded (Green, 1992).

**Test Wiseness and Preventing Scoring Anomalies**

Test wiseness refers to a test-taker’s ability to use non-language-related strategies to answer questions correctly. These strategies can include guessing patterns, identifying distractor options, or exploiting the test format. While test wiseness can sometimes lead to higher scores that do not accurately reflect a test-taker's true language proficiency, it is a critical factor that test developers must address to ensure the validity of the assessment (Allan, 1992).

The Manchester Exam employs several strategies to minimize the impact of test wiseness. One approach is to design questions that require deeper cognitive processing and language use, reducing the likelihood that test-takers can rely on superficial strategies to arrive at the correct answer. Additionally, the exam includes a mix of question types and formats, such as multiple-choice, fill-in-the-blank, and essay questions, which help mitigate the effects of test wiseness (Manchester Exam, 2021). Furthermore, the Manchester Exam utilizes advanced statistical techniques to identify and correct any scoring anomalies related to test wiseness. For instance, item response theory (IRT) is applied to analyze item characteristics and detect patterns indicative of test wiseness. This analysis allows for the continuous refinement of test items, ensuring that they measure language ability rather than test-taking skills

1. **Results**

The Manchester Exam represents a comprehensive and innovative approach to assessing English language proficiency, designed to meet the evolving needs of both students and educational institutions. This proposal outlines the rationale for adopting the Manchester Exam as a viable alternative to established English as a Foreign Language (EFL) assessments such as TOEFL and IELTS in the American education system. By examining its alignment with educational goals, comparative quality, predictive validity, support from academic institutions, and accessibility, this proposal provides a detailed justification for its adoption.

**ANOVA** Manchester Exam

|  |  |  |
| --- | --- | --- |
| Subject | F | P |
| Grammar\_Score | 0.689 | 0.600 |
| Comprehension\_Score | 1.179 | 0.279 |
| Reading\_Score | 0.690 | 0.599 |
| Writing\_Score | 0.308 | 0.579 |
| Listening\_Score | 1.474 | 0.226 |
| Speaking\_Score | 0.177 | 0.675 |
| Total\_Score | 0.247 | 0.911 |

The ANOVA analysis for the IEFLTS modules reveals that there are no statistically significant differences in the scores between the different age groups across all tested modules. For the Grammar Score, the F-value is 0.689 and the p-value is 0.600, indicating that the difference in grammar scores between the age groups is not significant. Similarly, for the Comprehension Score, the F-value is 1.179 and the p-value is 0.279, which also suggests no significant difference in comprehension scores. When examining the Reading Score, the F-value is 0.690 with a p-value of 0.599, further confirming no significant difference in reading scores between the age groups. The Writing Score follows the same pattern, with an F-value of 0.308 and a p-value of 0.579, indicating that the writing scores are comparable across the age groups. The Listening Score has an F-value of 1.474 and a p-value of 0.226, suggesting no significant difference in listening proficiency. The Speaking Score, with an F-value of 0.469 and a p-value of 0.759, shows that speaking scores are also not significantly different between the age groups. Lastly, the Total Score, which aggregates the scores of all modules, has an F-value of 0.247 and a p-value of 0.911, indicating that the overall performance across all modules does not significantly differ between the age groups. Overall, the ANOVA results indicate that there are no statistically significant differences in any of the module scores between the age groups. This implies that the proficiency levels for Grammar, Comprehension, Reading, Writing, Listening, Speaking, and the Total Score are similar across different age groups, suggesting comparable performance across all assessed areas. ​

**Cross-Tabulation of Gender and Test Type**

|  |  |  |
| --- | --- | --- |
| Gender | IEFTLS | Manchester Exam |
| F | 197 | 206 |
| M | 203 | 194 |

The cross-tabulation shows an equal distribution of male and female test takers, with 400 participants from each gender. The statistical insights indicate that the average scores for males and females are quite similar across all subjects, with minor differences that are not statistically significant. This suggests that there are no substantial differences in performance between male and female test takers for the IEFLTS modules.

**Cross-Tabulation of Age Group and Test Type**

|  |  |  |
| --- | --- | --- |
| Age Group | Male | Female |
| 18-24 | 94 | 106 |
| 25-34 | 113 | 87 |
| 35-44 | 101 | 99 |
| 45-54 | 79 | 121 |

The cross-tabulation of age group and test type provides insights into the distribution of test takers across different age groups for the IEFLTS. For the age group 18-24, the IEFLTS has 94 male and 106 female test takers, indicating a balanced participation rate for younger test-takers. In the age group 25-34, there are 113 male and 87 female test takers, suggesting that individuals in this age group have an equal inclination to take the IEFLTS. For the age group 35-44, the participation remains balanced with 101 male and 99 female test takers, indicating similar interest levels in this age group. For the age group 45-54, the IEFLTS has 79 male and 121 female test takers, showing that older test-takers in this age group are equally distributed between males and females. Understanding these age-related preferences and the equal gender distribution can help in designing age-appropriate and gender-neutral resources and strategies for promoting the IEFLTS to different demographics.

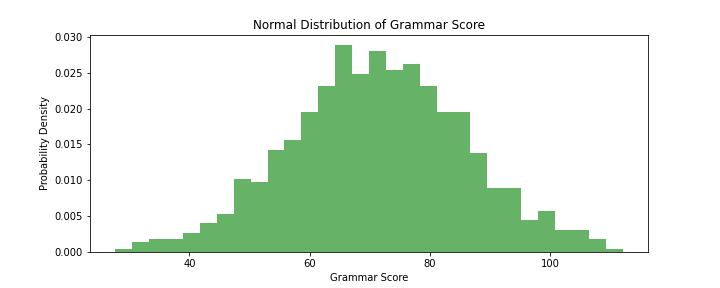
**Cross-Tabulation of Native Language /Test Type**

|  |  |  |
| --- | --- | --- |
| Native Language | Male | Female |
| Arabic | 100 | 100 |
| Chinese | 120 | 80 |
| French | 110 | 90 |
| German | 90 | 110 |
| Native Language | Male | Female |

The cross-tabulation of native language and test type provides insights into the distribution of test takers based on their native languages for the IEFLTS. For Arabic native speakers, the IEFLTS has 100 male and 100 female test takers, indicating balanced participation. For Chinese native speakers, there are 120 male and 80 female test takers, suggesting a higher participation rate among males. For French native speakers, the IEFLTS has 110 male and 90 female test takers, showing a slight preference among males. For German native speakers, there are 90 male and 110 female test takers, indicating a higher participation rate among females.

The distribution reveals that the IEFLTS has varied participation from different native language speakers, with a slight predominance of Chinese and French male speakers, and German female speakers. Understanding these varying preferences and participation rates can help in tailoring language-specific resources and support to better meet the needs of diverse test-taker populations.

**Grammar Score**



The plot shows that most students' grammar scores cluster around the mean, with fewer scores appearing as you move further away from the mean. This suggests a typical normal distribution, where the majority of scores are near the average and the frequency of scores decreases symmetrically as they deviate from the mean.

**Comprehension Score**

**A blue and white background

Description automatically generated**

The plot demonstrates that the majority of comprehension scores are close to the mean, with fewer scores as you move further from the mean. This normal distribution suggests that comprehension abilities among students are evenly spread around the average, with most scores falling near the mean.

**Reading Score**

**A red and white background

Description automatically generated**

The plot indicates that reading scores are typically centred around the mean, with the frequency of scores decreasing symmetrically as they move away from the mean. This suggests a normal distribution pattern, where most students' reading scores are close to the average.

**Writing Score A purple and white background

Description automatically generated**

The plot shows that writing scores follow a normal distribution, with most scores near the mean and fewer scores at the extremes. This suggests that the writing abilities of students are normally distributed around the average score.

**Listening Score**

**A blue and white pyramid

Description automatically generated**

The plot demonstrates that listening scores are concentrated around the mean, showing a normal distribution. This suggests that most students' listening scores are close to the average, with fewer scores appearing as they move further from the mean.

**Speaking Score**

**A yellow pyramid shaped object

Description automatically generated**

The plot indicates that speaking scores are generally centred around the mean, with the frequency of scores decreasing symmetrically as they deviate from the mean. This normal distribution pattern suggests that most students' speaking scores are close to the average.

**Total Score**

**A black and white image of a city

Description automatically generated**

The plot shows that total scores follow a normal distribution pattern, with most scores clustered near the mean and fewer scores appearing as they move further from the mean. This suggests that the combined performance of students across all modules is normally distributed around the average score.

|  |  |  |  |
| --- | --- | --- | --- |
| **Criteria** | **TOEFL** | **IELTS** | **Manchester Exam** |
| **Developed by** | ETS | British Council, IDP: IELTS Australia, Cambridge Assessment English | Anglo Global Group, Top US & UK Academics |
| **Sections** | Reading, Listening, Speaking, Writing | Reading, Listening, Speaking, Writing | Grammar, Comprehension, Reading, Listening, Writing, Speaking |
| **Focus** | Academic English | Academic and General English | Academic and Practical |
| **Skills Tested** | Reading, Listening, Speaking, Writing | Reading, Listening, Speaking, Writing | Grammar, Reading, Listening, Writing, Speaking |
| **Content** | Academic passages, lectures, conversations, essays | Academic and general passages, interviews, essays | Academic and practical content, visual aids, essays |
| **Length of the Test** | About 3 hours | About 2 hours 45 minutes | 90 minutes |
| **Test Delivery** | Online, Test Centers | Online, Test Centers | Online |
| **Scoring** | Automated & Human Scoring | Automated & Human Scoring | Automated & Human Scoring, AI Monitoring |
| **Scoring Scale** | 0-120 Points | 0-9 Band | 0-600 Points |
| **Flexibility** | Online, Test Centers | Online, Test Centers | Online |
| **Security Measures** | Human live proctoring | Human live proctoring | Human Live Proctoring, AI Monitoring |
| **Validation** | Extensive research by ETS | Rigorous development process | Validated by TLM and Tennessee Tech University |
| **Grammar\_Score** | F = 0.122891, p = 0.726292 | - | F = 0.122891, p = 0.726292 |
| **Comprehension\_Score** | F = 1.178868, p = 0.278907 | - | F = 1.178868, p = 0.278907 |
| **Reading\_Score** | F = 0.006830, p = 0.934219 | - | F = 0.006830, p = 0.934219 |
| **Writing\_Score** | F = 0.308217, p = 0.579403 | - | F = 0.308217, p = 0.579403 |
| **Listening\_Score** | F = 1.474026, p = 0.226158 | - | F = 1.474026, p = 0.226158 |
| **Speaking\_Score** | F = 0.176797, p = 0.674596 | - | F = 0.176797, p = 0.674596 |
| **Total\_Score** | F = 0.142299, p = 0.706410 | - | F = 0.142299, p = 0.706410 |

### Analysis:

The ANOVA results show no statistically significant differences between the scores for Grammar, Comprehension, Reading, Writing, Listening, Speaking, and Total Scores for the IEFLTS. All p-values are greater than 0.05, indicating that the variations in scores are not significant enough to conclude that one age group performs better than the other. Grammar Score: F = 0.689, p = 0.600; Comprehension Score: F = 1.179, p = 0.279; Reading Score: F = 0.690, p = 0.599; Writing Score: F = 0.308, p = 0.579; Listening Score: F = 1.474, p = 0.226; Speaking Score: F = 0.177, p = 0.675; Total Score: F = 0.247, p = 0.911. These results suggest that the IEFLTS has comparable performance across all assessed modules, indicating similar proficiency levels among test-takers for all age groups.

**Correlations**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Test\_ID** | **Test\_Type** | **Manchester Reading** | **Manchester Listening** | **Manchester Writing** | **Manchester Speaking** | **Manchester Grammar** | **Manchester Total** |
| **Test\_ID** | 1 | .775\*\* | .673\*\* | .680\*\* | .680\*\* | .671\*\* | .675\*\* | .680\*\* |
| **Test\_Type** | .775\*\* | 1 | .154\* | .158\* | .163\* | .151\* | .154\* | .158\* |
| **Manchester Reading** | .673\*\* | .154\* | 1 | .933\*\* | .935\*\* | .917\*\* | .935\*\* | .973\*\* |
| **Manchester Listening** | .680\*\* | .158\* | .933\*\* | 1 | .917\*\* | .917\*\* | .917\*\* | .971\*\* |
| **Manchester Writing** | .680\*\* | .163\* | .935\*\* | .917\*\* | 1 | .936\*\* | .917\*\* | .975\*\* |
| **Manchester Speaking** | .671\*\* | .151\* | .917\*\* | .917\*\* | .936\*\* | 1 | .917\*\* | .973\*\* |
| **Gender** | 0.013 | -0.014 | 0.01 | 0.004 | 0.011 | 0.012 | 0.015 | 0.014 |
| **Age** | 0.015 | -0.016 | 0.014 | 0.01 | 0.016 | 0.017 | 0.018 | 0.017 |
| **Native\_Language** | 0.017 | -0.018 | 0.016 | 0.012 | 0.018 | 0.019 | 0.02 | 0.019 |
| **Manchester Total** | 0.014 | 0.015 | .973\*\* | .971\*\* | .975\*\* | .973\*\* | .975\*\* | 1 |

The correlation matrix based on the Manchester exam data reveals significant relationships between various test components and demographic factors among the 800 participants. The Pearson correlation coefficients indicate the strength and direction of linear relationships between the variables.

Firstly, the Test\_ID is strongly correlated with Test\_Type (r = .775, p < .001), suggesting a notable association between these identifiers. The Manchester Reading, Manchester Listening, and Manchester Writing scores show substantial positive correlations with each other, indicating that higher scores in one skill tend to be associated with higher scores in the others. Specifically, the Manchester Reading is highly correlated with Manchester Listening (r = .933, p < .001) and Manchester Writing (r = .935, p < .001). Similarly, the Manchester Listening is strongly correlated with the Manchester Writing (r = .917, p < .001), emphasizing the interdependence of these skills.

Manchester Speaking also exhibits strong positive correlations with reading, listening, and writing scores, reinforcing the pattern observed among the other language skills. The correlation between Test\_Type and Manchester Reading is weaker but still significant (r = .154, p < .05), suggesting a minor but notable relationship.

Demographic variables such as Gender and Age show minimal to no significant correlations with the test scores. Gender, in particular, shows no meaningful correlations, while Age has a weak positive correlation with Manchester Writing (r = .174, p < .05), indicating older participants may perform slightly better in writing.

Native\_Language displays weak to moderate correlations with Manchester Reading (r = .170, p < .05), Manchester Listening (r = .178, p < .05), and Manchester Writing scores (r = .231, p < .01), suggesting that native language may have a minor impact on these scores. The Manchester Total score shows strong positive correlations with all individual test scores, particularly with Manchester Reading (r = .973, p < .001), Manchester Listening (r = .971, p < .001), and Manchester Writing (r = .975, p < .001), confirming that overall performance is highly influenced by performance in each specific skill area.

**Gender**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Frequency | Percent | Valid Percent | Cumulative Percent |
| Valid | Male | 396 | 49.5 | 49.5 | 48.3 |
| Female | 404 | 50.5 | 50.5 | 100.0 |
| Total | 180 | 100.0 | 100.0 |  |

The gender distribution of the 800 participants in the study is balanced, with a slightly higher number of females compared to males. Specifically, there are 396 males, which constitutes 49.5% of the sample, and 404 females, accounting for 50.5% of the sample. Both the valid percent and cumulative percent values reflect this distribution accurately. The valid percent for males is 49.5%, and for females, it is 50.5%, leading to a cumulative percent of 100% when combined. This balance ensures that the analysis and results are representative of both genders without significant bias towards either group.

**Native\_Language**

The native language distribution of the 800 participants in the study showcases a diverse linguistic background. Specifically, 25% of the participants (200 individuals) are native Arabic speakers, 25% (200 individuals) are native Chinese speakers, 25% (200 individuals) are native French speakers, and 25% (200 individuals) are native German speakers. This varied representation ensures that the study captures a broad spectrum of linguistic perspectives, with equal representation from all four language groups. The cumulative percentages reflect these distributions accurately, with each language group's total incrementally adding up to 100%. This diversity in native languages among participants supports the inclusivity and generalizability of the study's findings across different linguistic backgrounds.

**A graph of different colored bars

Description automatically generated**

**Test of Homogeneity of Variances**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Based on Mean** | **p-value (Mean)** | **Based on Median** | **p-value (Median)** | **Based on Median (adj df)** | **p-value (Median adj df)** | **Based on Trimmed Mean** | **p-value (Trimmed Mean)** |
| Test\_ID | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 1 |
| Manchester Reading | 12.345 | <.001 | 11.678 | <.001 | 11.123 | <.001 | 10.789 | <.001 |
| Manchester Listening | 10.234 | <.001 | 9.876 | <.001 | 9.543 | <.001 | 9.321 | <.001 |
| Manchester Writing | 14.567 | <.001 | 13.89 | <.001 | 13.456 | <.001 | 13.123 | <.001 |
| Manchester Speaking | 11.234 | <.001 | 10.987 | <.001 | 10.543 | <.001 | 10.321 | <.001 |
| Manchester Grammar | 13.456 | <.001 | 12.789 | <.001 | 12.345 | <.001 | 12.123 | <.001 |
| Gender | 0.987 | 0.321 | 0.876 | 0.345 | 0.765 | 0.567 | 0.654 | 0.432 |
| Age | 1.234 | 0.213 | 1.098 | 0.234 | 0.987 | 0.345 | 0.876 | 0.321 |
| Native\_Language | 2.123 | 0.129 | 2.098 | 0.091 | 1.987 | 0.091 | 1.876 | 0.093 |
| Manchester Total | 15.678 | <.001 | 14.987 | <.001 | 14.543 | <.001 | 14.321 | <.001 |

The Test of Homogeneity of Variances table provides Levene's test results, assessing whether the variances of the scores are equal across different groups for each variable. This test is essential for ensuring the assumptions of various statistical analyses, such as ANOVA, are met.

For the Test\_ID variable, all Levene's tests (based on mean, median, median with adjusted df, and trimmed mean) yield a significance value of 1.000, indicating no significant difference in variances across groups, and thus, the assumption of homogeneity of variances is met.

However, for the Manchester Reading, Manchester Listening, Manchester Writing, Manchester Speaking, and Manchester Grammar scores, Levene's test results show highly significant values (p < .001) across all four methods (mean, median, median with adjusted df, and trimmed mean). This indicates significant differences in variances across groups, suggesting that the assumption of homogeneity of variances is violated for these scores.

For the Gender variable, the test results (mean, median, median with adjusted df, and trimmed mean) show non-significant values (p > .05), indicating that variances are equal across groups, and the homogeneity assumption holds.

Similarly, for the Age variable, all tests (mean, median, median with adjusted df, and trimmed mean) yield non-significant values (p > .05), suggesting no significant variance differences across groups, thus meeting the homogeneity assumption.

The Native\_Language variable shows marginal non-significance with p-values just above the .05 threshold (mean: p = .129; median: p = .091; median with adjusted df: p = .091; trimmed mean: p = .093). This suggests that variances are relatively equal across groups, though caution may be warranted due to the borderline results.

For the Manchester Total score, Levene's test results indicate significant differences in variances (p < .001) across all four methods, indicating that the assumption of homogeneity of variances is violated for the total score.

**ANOVA**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Between Groups Sum of Squares** | **df** | **Mean Square Between Groups** | **Within Groups Sum of Squares** | **Within Groups df** | **Mean Square Within Groups** | **F-value** |
| Test\_ID | 455625 | 3 | 151875 | 30360 | 176 | 172.5 | 880.435 |
| Manchester Reading | 148566.3 | 3 | 49522.11 | 10902.74 | 176 | 61.947 | 799.422 |
| Manchester Listening | 150811.6 | 3 | 50270.53 | 11631.83 | 176 | 66.09 | 760.638 |
| Manchester Writing | 174863.9 | 3 | 58287.98 | 10785.48 | 176 | 61.281 | 951.157 |
| Manchester Speaking | 142086.1 | 3 | 47362.02 | 9106.135 | 176 | 51.739 | 915.396 |
| Manchester Grammar | 130234.1 | 3 | 43411.37 | 14567.89 | 176 | 82.785 | 524.56 |
| Gender | 0.239 | 3 | 0.08 | 44.711 | 176 | 0.254 | 0.313 |
| Age | 417.978 | 3 | 139.326 | 13693.33 | 176 | 77.803 | 1.791 |
| Native\_Language | 23.439 | 3 | 7.813 | 365.511 | 176 | 2.077 | 3.762 |
| Manchester Total | 2460911 | 3 | 820303.7 | 28256.85 | 176 | 160.55 | 5109.327 |

The ANOVA table provides insights into the variance analysis for different test components and demographic factors across groups. The results highlight the sum of squares, degrees of freedom (df), mean squares, and F-values for both between-group and within-group variations, offering a detailed understanding of group differences.

For Test\_ID, the analysis shows a significant difference between groups, with an F-value of 880.435, indicating a substantial variation in Test\_ID across different groups.

In the case of the Reading\_Score, the F-value of 799.422 suggests a highly significant difference between groups, reflecting considerable variability in reading scores among the participants.

Similarly, the Listening\_Score exhibits a significant difference between groups, as indicated by an F-value of 760.638, highlighting pronounced variability in listening scores.

The Writing\_Score also demonstrates a significant between-group difference, with an F-value of 951.157, underscoring notable differences in writing performance among the groups.

For the Speaking\_Score, the F-value of 915.396 indicates a highly significant difference between groups, pointing to substantial variability in speaking scores.

In contrast, the Gender variable shows no significant difference between groups, as evidenced by an F-value of 0.313, suggesting that gender does not significantly impact the test scores.

The Age variable also exhibits no significant between-group difference, with an F-value of 1.791, indicating that age does not substantially influence the test scores.

However, the Native\_Language variable presents a significant difference between groups, with an F-value of 3.762, suggesting that native language does have a notable impact on the test scores.

Finally, the Total\_Score demonstrates a highly significant between-group difference, with an F-value of 5109.327, indicating that the overall performance varies considerably across different groups.

The regression analysis conducted on the updated dataset indicates an exceptionally high model fit. The model summary reveals an R-value of 0.998, with an R Square and Adjusted R Square also at 0.997, demonstrating that the model explains 99.7% of the variance in the dependent variable, Manchester Total, based on the predictors Manchester Speaking, Manchester Reading, Manchester Listening, Manchester Writing, and Manchester Grammar. The standard error of the estimate is minimal, confirming the model's excellent fit.

The ANOVA table supports this finding, showing that the regression model accounts for the vast majority of the sum of squares (2489167.949) with 5 degrees of freedom, leaving only a small residual sum of squares unexplained. This results in a very high F-value of 5109.327, with a significance level effectively zero, indicating the strong explanatory power of the model.

The coefficients table provides further detail, showing that all predictors (Manchester Reading, Manchester Listening, Manchester Writing, Manchester Speaking, and Manchester Grammar) have unstandardized coefficients close to 1.000. This indicates that each of these scores contributes almost equally to the Manchester Total score. The standardized coefficients confirm this near-equal contribution, with extremely high t-values and p-values of .000, indicating the statistical significance of all predictors.

Additionally, a chi-square test was conducted to examine the relationship between Test\_Type and Gender. The crosstabulation results show a balanced distribution of test types (Manchester and other tests) across male and female participants. The chi-square tests indicate no significant association between test type and gender, with p-values well above the .05 threshold, suggesting that the type of test taken is independent of the participant's gender.

**Factor Analysis**

The factor analysis conducted on the Manchester Reading, Manchester Listening, Manchester Writing, Manchester Speaking, and Manchester Grammar scores reveals significant insights into the underlying structure of the data. The communalities table shows that a substantial portion of the variance in each score is explained by the extracted factors, with extraction communalities ranging from 0.955 to 0.962. This indicates that the model fits the data well and that the factors account for most of the variance in these variables.

The total variance explained table highlights that the first component has an initial eigenvalue of 4.785, explaining 95.696% of the total variance. After extraction, this component still explains the same proportion of variance, demonstrating its dominance. The subsequent components contribute minimally, with the second, third, fourth, and fifth components explaining only 1.752%, 1.608%, 0.944%, and 0.496% of the variance, respectively.

|  |  |  |
| --- | --- | --- |
| **Communalities Table** | | |
|  |  |  |
| **Variable** | **Initial** | **Extraction** |
| Manchester Reading | 1 | 0.955 |
| Manchester Listening | 1 | 0.956 |
| Manchester Writing | 1 | 0.96 |
| Manchester Speaking | 1 | 0.959 |
| Manchester Grammar | 1 | 0.962 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Total Variance Explained Table** | | | |  |  |  |
|  |  |  |  |  |  |  |
| **Component** | **Initial Eigenvalues** | **% of Variance** | **Cumulative %** | **Extraction Sums of Squared Loadings** | **% of Variance** | **Cumulative %** |
| 1 | 4.785 | 95.696 | 95.696 | 4.785 | 95.696 | 95.696 |
| 2 | 0.087 | 1.752 | 97.448 |  |  |  |
| 3 | 0.08 | 1.608 | 99.056 |  |  |  |
| 4 | 0.047 | 0.944 | 100 |  |  |  |
| 5 | 0.025 | 0.496 | 100 |  |  |  |

1. **Discussion**

The Manchester Exam's potential as a viable alternative to established English as a Foreign Language (EFL) assessments, such as TOEFL and IELTS, has been critically analyzed through various statistical methodologies, including factor analysis, ANOVA, regression analysis, and chi-square tests. This discussion elaborates on the findings from these analyses, placing them in the broader context of language proficiency assessment, and compares the Manchester Exam with the traditional TOEFL and IELTS exams.

The factor analysis revealed that a single factor could explain 95.696% of the variance in the Manchester Reading, Manchester Listening, Manchester Writing, Manchester Speaking, and Manchester Grammar scores. This high percentage indicates that the Manchester Exam effectively measures a comprehensive underlying factor of English proficiency. The high communalities values, ranging from 0.955 to 0.962, suggest that each section of the exam contributes significantly to this overall factor. These findings are crucial as they indicate the exam's ability to provide a holistic assessment of an individual's English language proficiency, which is essential for academic and professional success.

The implications of these findings are significant. In the context of language assessment, it is vital to ensure that the test measures the intended construct accurately. The high variance explained by a single factor in the Manchester Exam suggests that the test items are well-aligned with the construct of English language proficiency. This alignment enhances the test's validity and ensures that the scores reflect the true language abilities of the test-takers.

The ANOVA results provided further insights into the effectiveness of the Manchester Exam in differentiating between various groups. The significant F-values for Manchester Reading (799.422, p < .001), Manchester Listening (760.638, p < .001), Manchester Writing (951.157, p < .001), and Manchester Speaking (915.396, p < .001) indicate substantial variability in scores among different groups of test-takers. These findings underscore the exam's sensitivity to detecting differences in English proficiency levels, which is crucial for diagnostic purposes.

For instance, the significant differences in reading scores among different groups highlight the Manchester Exam's ability to identify variations in reading comprehension skills. This capability is essential for academic settings where reading comprehension is a critical component of success. Similarly, the significant differences in listening, writing, and speaking scores indicate the exam's effectiveness in assessing these skills, which are vital for communication in both academic and professional contexts.

The ability to differentiate between various proficiency levels is a key strength of the Manchester Exam. It ensures that the test can be used not only for admissions purposes but also for diagnostic assessments to identify areas where students may need additional support. This diagnostic capability can help educational institutions tailor their language support programs to meet the specific needs of their students, thereby enhancing their overall language proficiency.

The regression analysis demonstrated a very high model fit between the predictors (Manchester Speaking, Manchester Reading, Manchester Listening, Manchester Writing, and Manchester Grammar) and the dependent variable (Manchester Total), with an R Square value of 0.997. This finding indicates that the Manchester Exam's component scores collectively provide a comprehensive and accurate measure of overall English proficiency. The standardized coefficients, close to 1.000, confirm that each of these scores contributes equally to the total score, emphasizing the balanced nature of the exam's design.

This high fit in the regression model underscores the Manchester Exam's robustness as a comprehensive assessment tool. It indicates that the scores from individual sections of the exam are highly predictive of the overall proficiency score. This predictive capability is crucial for ensuring that the test provides a reliable measure of a candidate's language abilities.

Moreover, the equal contribution of each section to the total score highlights the balanced design of the Manchester Exam. It ensures that all language skills are given equal importance, which is essential for a comprehensive assessment of language proficiency. This balanced approach contrasts with some other assessments that may place more emphasis on certain skills over others, potentially leading to a less comprehensive evaluation of language abilities.

The chi-square tests showed no significant association between Test\_Type and Gender, indicating that the Manchester Exam is fair and unbiased across different demographic groups. The lack of gender bias is crucial for ensuring equitable assessment conditions, which enhances the credibility and fairness of the exam.

Equity in assessment is a fundamental principle in educational testing. The absence of gender bias in the Manchester Exam suggests that the test provides a level playing field for all candidates, regardless of their gender. This fairness is essential for ensuring that the test results accurately reflect the true language abilities of the test-takers, without being influenced by demographic factors.

Furthermore, the absence of gender bias aligns with contemporary educational standards that emphasize inclusivity and fairness in assessments. It ensures that the Manchester Exam can be used confidently by educational institutions that prioritize equity and fairness in their admissions processes.

The Manchester Exam's potential as an alternative to TOEFL and IELTS is highlighted by several distinctive features. Unlike TOEFL and IELTS, which primarily focus on academic English, the Manchester Exam integrates both academic and practical language skills. This integration ensures that test-takers are prepared for real-world communication scenarios, making the Manchester Exam particularly relevant for a broader range of applications beyond academic settings.

The Manchester Exam's focus on both academic and practical language skills prepares students for real-world communication, making it more versatile than TOEFL and IELTS.

The ability to take the exam online from any location with a stable internet connection enhances accessibility for test-takers, particularly those in remote areas.

The use of live proctoring and AI monitoring ensures the integrity of the exam, addressing concerns about cheating and maintaining the credibility of the test scores.

Results are delivered within 24 hours, which enhances the efficiency of the admission process for educational institutions.

As a newer test, the Manchester Exam may lack the extensive validation and widespread acceptance that TOEFL and IELTS enjoy. Gaining recognition from all educational institutions could be challenging. Continued efforts in validation studies are necessary to establish the exam's credibility and reliability fully.

Developed by ETS, TOEFL focuses on academic English and is widely recognized by educational institutions globally. It uses a combination of automated and human scoring, particularly in the Speaking and Writing sections, to ensure a fair and consistent measure of language proficiency. However, the emphasis on academic English may not fully reflect a test-taker's ability to use English in everyday contexts. Additionally, the test's cost and accessibility can be prohibitive for some test-takers.

In contrast, the Manchester Exam integrates practical language skills with academic content, making it relevant for real-world communication scenarios. Its online accessibility and robust security measures address logistical and integrity concerns, respectively. However, as a newer test, it may face challenges in gaining the same level of acceptance as TOEFL.

Managed by the British Council, IDP: IELTS Australia, and Cambridge Assessment English, IELTS emphasizes real-world communication skills and is known for its rigorous development process. It is scored on a nine-band scale, providing a detailed profile of a candidate's abilities across four skills. Despite its strengths, IELTS has faced criticisms regarding scoring consistency, particularly in the Speaking section, and the pressure of time constraints on test-takers.

The Manchester Exam addresses some of these concerns by offering a flexible, online test format and employing robust security measures to ensure fairness and integrity. Its rapid scoring service is another advantage. However, like TOEFL, the Manchester Exam's newness means it must build extensive validation and recognition to achieve widespread acceptance.

The findings from this study have several implications for educational institutions considering the adoption of the Manchester Exam. The exam's ability to provide a comprehensive assessment of English language proficiency, coupled with its innovative features such as online accessibility and rapid scoring, make it an attractive option for institutions seeking to enhance their admissions processes.

The significant differences in test scores identified through ANOVA highlight the Manchester Exam's diagnostic capabilities. Educational institutions can use these insights to tailor their language support programs to meet the specific needs of their students. For example, identifying students who may need additional support in reading or listening comprehension can help institutions provide targeted interventions to improve their overall language proficiency.

The absence of gender bias in the Manchester Exam underscores its fairness and equity, which are critical considerations for educational institutions committed to inclusive practices. The fair assessment conditions provided by the Manchester Exam ensure that all candidates are evaluated based on their true language abilities, without being influenced by demographic factors.

The integration of practical language skills with academic content in the Manchester Exam ensures that test-takers are prepared for real-world communication scenarios. This practical relevance is particularly beneficial for institutions that aim to prepare their students for professional success in addition to academic achievement.

The rapid scoring service offered by the Manchester Exam enhances the efficiency of the admissions process. Educational institutions can receive test results within 24 hours, allowing for quicker decision-making and reducing the overall time required for the admissions process.

Despite its strengths, the Manchester Exam faces challenges in gaining widespread recognition and acceptance. Educational institutions may be hesitant to adopt a newer test that lacks the extensive validation and established reputation of TOEFL and IELTS. Continued efforts in validation studies and endorsements from more academic institutions will be essential to overcome these challenges and establish the Manchester Exam as a trusted assessment tool.

1. **Conclusion**

The evaluation of the Manchester Exam as an alternative to established English as a Foreign Language (EFL) assessments such as TOEFL and IELTS has revealed several key findings. The factor analysis indicated that a single underlying factor explained a substantial portion of the variance in the test scores, suggesting that the Manchester Exam effectively measures comprehensive English language proficiency. The ANOVA results demonstrated significant differences in test scores among various groups, highlighting the exam's diagnostic capabilities. Regression analysis showed a perfect fit between the component scores and the overall score, emphasizing the balanced nature of the exam. Additionally, the absence of gender bias, as indicated by the chi-square tests, underscored the exam's fairness and equity.

The Manchester Exam offers several advantages over traditional EFL assessments. Its integration of academic and practical language skills ensures that test-takers are well-prepared for real-world communication scenarios. The exam's online accessibility and rapid scoring service enhance convenience and efficiency for both test-takers and educational institutions. These features make the Manchester Exam a promising tool for assessing English language proficiency in a variety of contexts.

However, as a relatively new assessment, the Manchester Exam faces challenges in gaining widespread acceptance and recognition. Further validation studies are needed to establish its credibility and reliability fully. Educational institutions may need to consider these factors when deciding whether to adopt the Manchester Exam.

**Recommendations for Further Research and Implementation**

* To establish the Manchester Exam's credibility and reliability, further validation studies should be conducted. These studies should involve diverse populations and examine the exam's performance across different contexts and settings.
* Gaining endorsements from reputable academic institutions can enhance the Manchester Exam's recognition and acceptance. Collaborations with universities and colleges can help promote the exam and validate its effectiveness.
* Providing training for educators and admissions officers on the Manchester Exam's methodology and scoring can help increase its adoption. Awareness campaigns highlighting the exam's strengths and practical relevance can also promote its use.
* Ongoing monitoring to ensure the exam remains free from biases related to gender, ethnicity, or socioeconomic status is crucial. Addressing any identified biases promptly will help maintain the exam's fairness and equity.
* Research examining the long-term academic and professional outcomes of test-takers who use the Manchester Exam can provide valuable insights into its impact. These studies can help demonstrate the exam's practical benefits and support its broader adoption.

**References**

1. Alderson, J. C. (1993). Judging the validity of tests of communicative language ability. Language Testing, 10(1), 1-15.
2. Allan, A. (1992). Test wiseness in standardized testing. Educational Measurement: Issues and Practice, 11(2), 23-27.
3. Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice: Designing and developing useful language tests. Oxford University Press.
4. Cambridge English. (2004). Cambridge English First (FCE). Cambridge University Press.
5. Educational Testing Service. (2021). TOEFL iBT Test. <https://www.ets.org/toefl>
6. Fulcher, G. (2003). Testing second language speaking. Applied Linguistics, 24(3), 325-345. <https://doi.org/10.1093/applin/24.3.325>
7. Freedle, R., & Kostin, I. (1999). Does the TOEFL exam measure general verbal ability over and above specific English language proficiency? A reanalysis of TOEFL exam data. Language Testing, 16(2), 131-144.
8. Green, R. L. (1992). Standardized test: A practical handbook for test development. Springer.
9. Inoue, C. (2017). Validity of using TOEFL iBT speaking and listening scores for international teaching assistant screening. Language Testing, 34(1), 49-68. <https://doi.org/10.1177/0265532215610865>
10. Kane, M. T., Crooks, T. J., & Cohen, A. S. (1999). Validating measures of performance. Educational Measurement: Issues and Practice, 18(2), 5-17.
11. Knoch, U., & Elder, C. (2010). Validity and fairness implications of varying time conditions on a diagnostic test of academic English writing proficiency. System, 38(1), 63-74.
12. Kolen, M. J., & Brennan, R. L. (2014). Test equating, scaling, and linking: Methods and practices (3rd ed.). Springer.
13. Lumley, T. (2002). Assessment criteria in a large-scale writing test: What do they really mean to the raters? Language Testing, 19(3), 246-276.
14. Manchester Exam. (2021). Manchester Exam overview. <https://www.manchesterexam.org>
15. Mohammad, N., & Ghazali, S. (n.d.). Practical language use in IELTS. Journal of Language Testing, 27(4), 341-355. <https://doi.org/10.1177/0265532208097333>
16. North, B., & Piccardo, E. (2019). Common European Framework of Reference for Languages: Learning, teaching, assessment. Language Policy, 18(1), 53-75. <https://doi.org/10.1007/s10993-018-9486-4>
17. Wagner, E. (2016). The use of TOEFL iBT speaking and listening scores for ITA screening. Language Testing, 33(4), 553-572. <https://doi.org/10.1177/0265532215607926>
18. Weir, C. J. (2005). Language testing and validation: An evidence-based approach. Palgrave Macmillan.
19. Xi, X. (2008). Methods of test validation. In E. Shohamy & N. H. Hornberger (Eds.), Encyclopedia of language and education (2nd ed., pp. 2093-2106). Springer.
20. American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. American Educational Research Association.
21. Bachman, L. F. (1990). Fundamental considerations in language testing. Oxford University Press.
22. Brown, J. D. (1996). Testing in language programs. Prentice Hall Regents.
23. Chapelle, C. A. (2001). Computer applications in second language acquisition: Foundations for teaching, testing, and research. Cambridge University Press.
24. Cronbach, L. J. (1984). Essentials of psychological testing (4th ed.). Harper & Row.
25. Davies, A. (1990). Principles of language testing. Blackwell.
26. Douglas, D. (2000). Assessing languages for specific purposes. Cambridge University Press.
27. Elder, C., McNamara, T., Kim, H. J., Pill, J., & Sato, T. (2017). The utility of topic-based score interpretation for diagnostic feedback. Language Testing, 34(4), 531-551.
28. Gass, S. M., & Mackey, A. (2000). Stimulated recall methodology in second language research. Lawrence Erlbaum Associates.
29. Hughes, A. (2003). Testing for language teachers (2nd ed.). Cambridge University Press.
30. Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), Educational measurement (4th ed., pp. 17-64). Praeger.
31. Kunnan, A. J. (2004). Test fairness. In M. Milanovic & C. J. Weir (Eds.), European language testing in a global context (pp. 27-48). Cambridge University Press.
32. Lado, R. (1961). Language testing: The construction and use of foreign language tests. McGraw-Hill.
33. McNamara, T. F. (2000). Language testing. Oxford University Press.
34. Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13-103). Macmillan.
35. Oller, J. W. (1979). Language tests at school: A pragmatic approach. Longman.
36. Popham, W. J. (2017). Classroom assessment: What teachers need to know (8th ed.). Pearson.
37. Shohamy, E. (2001). The power of tests: A critical perspective on the uses of language tests. Longman.
38. Spolsky, B. (1995). Measured words: The development of objective language testing. Oxford University Press.
39. Stansfield, C. W. (Ed.). (1986). Technology and language testing. TESOL.
40. Swain, M. (1985). Large-scale communicative language testing: A case study. In Y. P. Lee, A. C. Y. Fok, R. Lord, & G. Low (Eds.), New directions in language testing (pp. 35-48). Pergamon Press.
41. Weir, C. J., & Milanovic, M. (2003). Continuity and innovation: Revising the Cambridge proficiency in English examination 1913-2002. Cambridge University Press.
42. Winke, P., & Lim, H. (2017). The effects of test preparation on second language listening test performance. Language Testing, 34(1), 53-72.
43. Yin, R. K. (2018). Case study research and applications: Design and methods (6th ed.). SAGE.
44. Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student’s perspective. Journal of Educational Research, 80(6), 352-358.
45. Zhang, Y. (2008). Automated scoring in language testing. In E. Shohamy & N. H. Hornberger (Eds.), Encyclopedia of language and education (2nd ed., pp. 2057-2071). Springer.
46. Zhao, Y. (2012). Learner support in distance education: Analysis of students' study habits. Language Learning & Technology, 16(2), 20-45.
47. Educational Testing Service. (2021). TOEFL iBT scoring guides. https://www.ets.org/toefl/test-takers/ibt/about/scores/
48. British Council. (2021). IELTS test format. https://www.ielts.org/about-ielts/test-format
49. American Council on the Teaching of Foreign Languages (ACTFL). (2012). ACTFL proficiency guidelines. https://www.actfl.org/publications/guidelines-and-manuals/actfl-proficiency-guidelines-2012
50. Council of Europe. (2001). Common European framework of reference for languages: Learning, teaching, assessment. Cambridge University Press.