Establishing Internal Consistency of English Language Proficiency Components of the English Placement Test (EPT)

Hairul Azhar Mohamad[*], Wardah Ismail, Muhammad Nasiruddin Aziz, Nasiha Nasrudin, Pavithran Ravinthra Nath, Amir Lukman Abd Rahman, Muhammad Haziq Abd Rashid, Hadayat Rahmah Hasan

Academy Pengajian Bahasa, University Technology MARA, Shah Alam, Malaysia

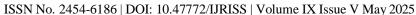***Corresponding Author**

ABSTRACT

This study aims to analyse the internal consistency of English Placement Tests (EPTs) emphasising its reliability and validity in measuring Malaysian students' English proficiency level. The employment of EPT which has been aligned with the Common European Framework of Reference (CEFR) is crucial in identifying language courses that may aid students' academic success. As this study utilises quantitative research design through the application of secondary research approach, correlational methods were used as an effective means to evaluate the internal consistency of listening, reading and grammar aspects in the database of EPT test papers of 5,423 diploma students. Results show a strong internal consistency in which significant positive correlations are found between the first and second half scores within all aspects of the EPT test. Hence, implying the effectiveness of EPTs in assessing students' proficiency level in accordance with the Common European Framework of Reference (CEFR). Aside from that, this study sheds light on the significance of internal consistency as to ensure the quality control in language assessment. Therefore, it is recommended that future research addresses other possible contributing factors to the quality of the test paper development by continuously refining the quality control approaches in assessing students' English proficiency. Overall, this study underscores that EPT is deemed to be reliable in measuring language proficiency among students alongside cultivating continuous improvement in accordance with international standards to optimise students' academic results.

Keywords: English Placement Test (EPT), listening, reading, grammar, English proficiency

INTRODUCTION

Background of Study

In general terms, consistency aspects of tests can be observed through the two (2) factors of (i) reliability and (ii) internal consistency between test items (which measure the same construct). Internal consistency, as argued by many scholars, is essential to ensure the overall reliability and validity of a test (Fulcher, 1997; Chung et al., 2015; Sheerah & Yadav, 2022; Sood, 2017). This is especially the case for tests with specific and significant purposes. One example is the English Placement Tests (EPTs) which are meant to assess students' English proficiency before placing them in the language courses that meet their learning needs. It is important as the students will receive the necessary and suitable support according to their English level competence for their academic success. This assertion is maintained by Sood (2017) who suggested that EPTs help to group students according to their language proficiency so that the appropriate and effective language instructions can be administered. Mohamad et al. (2025) further underscores this argument by highlighting the requirement of the English placement tests at the university levels to have a balanced approach towards listening and speaking as well as grammar and reading. This approach assists the university to measure students' language proficiency in a holistic manner so that the students' final scores can be utilised for the students' placement into the relevant university programs effectively.

Having shown the consensus on how important EPTs are, the developments of any types of English placement test as an entry form of evaluation in recent years have been observed to result in arising challenges. This includes the use of standardised tests like Test of English as a Foreign Language (TOEFL) and International English Language Testing System (IELTS) which have been validated for their reliability and internal consistency, making them dependable tools for student placement purposes (Bakri, 2022). More recently, the world has seen a linguistic standard guideline that is considerably growing in popularity - the Common European Framework of Reference for Languages (CEFR). Since then, the guideline has resulted in many changes in language testing globally as many countries have been aligning their assessments to the standard (Mohd Ali et al., 2018). Despite its popularity, the framework has been argued to lack connection with stakeholders, socio-educational contexts, as well as empirical validation, particularly in non-European contexts. This, among others, includes Malaysian EPTs.

In the Malaysian context, ongoing efforts are currently being made to align English language tests with CEFR, in line with the English Language Roadmap 2015-2025 (Mohd Ali et al., 2018). The widely used standardised test for Malaysian university admission and placement purposes is the Malaysian University English Test (MUET) (Baharum et al., 2021; Rethinasamy & Chuah, 2012). Its reliability has been proven by many studies which revealed positive correlations between students' academic achievement (particularly in the reading and speaking components) and the test (Baharum et al., 2021).

Though the standardised tests mentioned above have mostly been validated, there have been concerns about the quality control procedures upon practising placement tests in some contexts. For instance, a study in Chinese higher education raised concerns about the reliability, validity, and overall usefulness of English placement tests, due to the lack of quality control procedures (Fan & Jin, 2019). This, therefore, suggests that continuous efforts must be made to ensure the validity, reliability, and alignment (with international standards) of all assessment tests in Malaysia. This is especially true for EPTs which have recently gone through rigorous processes of alignment efforts with the CEFR guidelines, indicating an urgent need to examine its internal consistency. This situation, therefore, presents a research avenue that is worth exploring to maintain the effectiveness of EPTs in the Malaysian context - this is exactly what the present study intends to do.

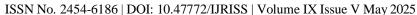### Objective of the Study and Research Questions (RQs)

EPTs evaluate students' several proficiency components. After its alignment efforts with CEFR, the research aimed to examine the internal consistency of an EPT of a well-established public university in Malaysia (or UiTM), herein referred to as EPT-PUiTM. Hence, the following research questions were outlined:

1. Does the Listening component of the EPT-PUiTM contain internal consistency?

2. Does the Reading component of the EPT- PUiTM contain internal consistency?

3. Does the Grammar component of the EPT- PUiTM contain internal consistency?

LITERATURE REVIEW

Theoretical Framework

Many studies have been conducted to determine the reliability and the internal consistency of both CEFR-aligned and non-CEFR tests. Through these research endeavours, they have used different approaches and have thus provided useful information to the field of language testing and evaluation. Studies on the CEFR-aligned tests have mainly focused on the development and the validation of tests focusing on all the language abilities. For example, based on with Classical Test Theory (CTT) and The Test of English for International Communication (TOEIC) scores, Waluyo et al. (2024) developed a CEFR-type test for undergraduates with the help of CTT to test listening and reading multiple-choice tests on 2,248 first year and 3,655 first and second-year students. The results showed that item difficulty and discrimination indices were favourable while the reliability coefficients were also high and computed using Cronbach's alpha, Kuder-Richardson formulas, and split-half reliability coefficients. Additionally, the test had a high predictive validity on TOEIC scores, which

may help in the formation of university-level English proficiency tests that integrate CEFR levels with CTT analysis.

Sridhanyarat et al. (2021) developed the Silpakorn Test of English Proficiency (STEP) and aligned it with CEFR. Thorough analysis of its validity and reliability also showed that the assessment was valid and reliable to be used by their university. Similarly, Hidri (2021) investigated the alignment process of International English Language Competency Assessment (IELCA) for the B1, B2, C1 and C2 levels onto CEFR using Many-Facet Rasch Model (FACETS_MFRM) application software. FACETS analysis is a powerful tool for understanding the complexities of rating processes and identifying areas for improvement, according to Issues in Language Teaching. This study also applied the Council of Europe's five linking stages to check the validity of the mapping exercise. Teachers' estimates showed the level of agreement which ranged from 74.4 per cent to 99.34 per cent, while FACETS analysis revealed a good global model fit and high reliability of the judgment process after the training of the raters.

While these studies highlighted the strengths of CEFR-aligned assessments, they also identified some key areas of challenges. Harsch and Hartig (2015) highlighted the variability of the judges' understanding of the CEFR levels; the judges used different criteria and descriptors, which raised a concern about the validity of the levels and the interpretation of the scores. Similarly, Leung and Jenkins (2020) claimed that CEFR frameworks are inadequate for identifying some important aspects of language practices, for instance, emotional intelligence. Therefore, this underscores the need to avoid rigid models of language proficiency so that more flexible models are used to better capture the actions that individuals engage in during their interactions in their everyday lives.

In contrast to CEFR-aligned assessments, the studies on non-CEFR assessments have also raised concerns on the use of traditional reliabilities. For instance, the Rosenzweig Picture-Frustration (P-F) study (Bernard, 1949), which explored projective and semi-projective techniques, argued that measures like analysis of variance and split-half methods, which assume item homogeneity, may not be suitable for the study under their consideration. Thus, the retest reliability analysis showed a statistically significant consistency in the main scoring categories, which also stresses the importance of using different approaches to the reliability of such assessments.

Recent studies on the internal consistency of English listening, reading, and grammar assessments has also helped to expand the scope of reliability evaluation. Mclean et al. (2021) analysed the online platform vocableveltest.org that includes 98 high-frequency English words tested in meaning-recall format. From the analysis, the study showed high internal consistency (Cronbach's $\alpha = 0.868$) and 98% match with human-marked responses. This research addressed limitations of existing vocabulary tests, particularly the written receptive meaning-recognition format, which is prone to guessing errors (Mclean et al., 2021).

Liu et al., (2020) examined the reliability and validity of the English listening test for adults using a statistical analysis software application. The findings showed that the selected listening test in general, had acceptable reliability and validity with clear distinguishing traits, but high item difficulty. As a result, Liu et al. (2020) suggested improving the listening materials' communicative skills as well as the authenticity of the listening test. In a survey using simultaneous reading and listening, Moussa-Inaty et al. (2011) found that those solely exposed to reading performed better on listening tasks than those subjected to both reading and listening. Moussa-Inaty et al. (2011) stated in their study that some learners, when examined through the lens of cognitive load theory framework, may improve their listening skills more efficiently just through reading materials rather than when they read and listen concurrently. Meanwhile, in a separate study, Kim and Kim (2017) assessed the validity of a 40-item English Placement Test (EPT). The test was designed and developed for a university General Language Program (GELP). The results of the analysis revealed that the EPT had a strong reliability with a Cronbach's alpha of 0.898, although over half of the questions were identified as difficult items. Furthermore, the ability of EPT to distinguish the upper-level and lower-level students clearly proves that EPT is not only highly reliable, but also an effective instrument in measuring students' competency levels within the GELP.

Several other studies have focused on the internal consistency of grammar components in English placement

tests. The current research explores the relationship between grammar testing and other language skills, its use in placement decisions, and as an indicator of learner's language proficiency. In the study on the online Yes/No test in a placement tool, the grammar placement tests were found to be strongly correlated with the overall placement decisions, with a correlation coefficient of 0.8. These findings, as argued by Harrington and Carey (2009), imply that a student's grammatical knowledge is crucial in determining the level of their language proficiency. Additionally, the study examined the efficiency of this format for placement testing, highlighting its correlation with grammar tests, nevertheless, it may be less sensitive to variations in placement levels when the response time is taken into consideration. Another similar study also reviewed the consistency of measured accuracy in grammar knowledge tests and writing. While grammar tests are often used, they may not accurately measure an individual's ability in real-time communicative tasks such as writing. Ahangari and Barghi (2012), showed a relationship between the scores on a grammar test and actual language usage. The internal consistency of grammatical components in English placement tests is affected by factors such as the test item type, their correlation with other language skills, and the criteria of the evaluators. Although grammar tests are essential for placements, in some instances, their capacity may be restricted in exhibiting real-time language use, and in delivering a comprehensive diagnostic information.

The above studies generally emphasise considering various factors in assessing the internal consistency of English language tests. Their results illustrate that the automated marking systems can be dependable for vocabulary tests, whereas other forms of the tests should be based on its individual components which are determined by the authentic and relevant skills. Furthermore, the cognitive loads linked to different learning methods would influence the test performance, implying the need for a customised assessment and instruction in language learning for general placement and any other specific purposes.

It can be concluded that the reviewed literature highlights significant advancements in the reliability and validity of both CEFR-aligned and non-CEFR language assessments, underscoring the importance of tailored approaches to testing. While CEFR-aligned tests demonstrate effective alignment and predictive validity, challenges such as variability in judges' interpretations and the inadequacy of rigid frameworks persist. Conversely, non-CEFR assessments reveal the limitations of traditional reliability measures, suggesting the necessity for innovative evaluation methods that consider cognitive load and real-world language use, ultimately advocating for a more flexible and context-sensitive approach to language proficiency assessment.
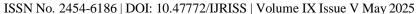
METHOD

Research Approach and Research Design

This study adopts a quantitative research approach, utilising secondary data research and correlational studies to evaluate the internal consistency and validity of an English placement test. Internal consistency is a crucial aspect of evaluating English placement tests. The validity of placement tests, such as TOEFL and IELTS, can be assessed using the Pearson Product Moment Correlation, with r-scores indicating the tests' reliability and validity across different language skills (reading, listening, speaking, and writing) (Bakri, 2022).

Sampling

The sample for this study consists of the English Placement Test questions and results extracted from the existing database of the teaching, learning, and testing platform of the university. Purposive sampling technique was employed in this research. This is because it is often employed in studies involving placement tests. Campbell et al. (2020), Zolkapli et al. (2024), and Zolkapli et al. (2025) highlights that purposive sampling improves the rigour of the study and trustworthiness of the data and results by better matching the sample to the research aims and objectives. This sampling method is particularly useful in qualitative research contexts, addressing aspects of credibility, transferability, dependability, and confirmability (Campbell et al., 2020). Despite its limited capacity in terms of generalisability, the research sought to ensure sampling representativeness from each faculty across all 22 UiTM branch campuses in Malaysia. This method follows a similar method employed by Mohamad et al. (2024) who applied the purposive sampling of students' Entrance English Competence test results in determining the students' levels of English proficiency competence. This method of purposive representative sampling was specifically done by allowing the EPT to be taken optionally

by diploma students who wanted to get the opportunity to be exempted from taking the first level diploma course in their first semester. Thus, the samples involved an approximate total of 40 to 50 percent of students from all faculties which offered diploma programs, who were in their first semester, and were randomly selected to participate in the English placement test. The sample size comprised a very large sample of 5,423 diploma students who had registered for their first semester.

Data Collection and Data Analysis

Data collection involved extracting marks from the database for each test taker, including individual question scores and overall section scores for Listening, Reading, and Grammar. The test consisted of 50 multiple-choice questions (MCQs) divided into three sections: 20 questions for Listening, 20 for Reading, and 10 for Grammar, with each question carrying 2 points, totalling 100 marks. For data analysis purposes, each section was split into halves for split reliability test purposes to facilitate a detailed examination of internal consistency using SPSS software (Field, 2013; Pallant, 2020). This approach ensures that the sample is not only representative but also large enough to provide statistically significant results, enhancing the reliability and generalisability of the findings. Split reliability tests were adopted in this analysis of the data because it is a method to measure the internal consistency or reliability of a test or measurement instrument. It tests how well the test items or scale scores are consistent with each other. Yusup (2018) mentions it as one of several testing techniques for internal consistency, depending on the type of instrument. Krus and Helmstadter (1987) presents the split-half coefficient of reliability as a conceptual precursor to modern formulations of internal-consistency reliability, facilitating understanding of internal-consistency reliability (Krus & Helmstadter, 1987; Yusup, 2018).

After obtaining the data for the halves or subsets of items, SPSS software was used to run Pearson Correlation Coefficients to determine the correlations between the two halves of each section. High correlations between the two halves would indicate that the sections are internally consistent and that the test is effectively measuring the intended competencies. The Pearson Correlation Coefficient is a robust statistical tool for this purpose, providing insights into the reliability of the test sections. The analysis aimed to demonstrate that the skills tested in each section were appropriately constructed to evaluate the students' competence in English. This methodological approach underscores the importance of rigorous statistical analysis in educational assessment, contributing to the development of reliable and valid testing instruments (Tabachnick & Fidell, 2019). Additionally, this method allows for the identification of any potential biases or inconsistencies within the test, ensuring that the assessment is fair and equitable for all students. It can be suggested that internal consistency measures, purposive sampling, and split-half reliability tests are valuable tools in evaluating English placement tests. However, their effective implementation varies across contexts, and there is a need for consistent quality control procedures to ensure the overall usefulness of placement tests in language assessment.

RESULTS

Internal Consistency of the Listening Part of the English Placement Test (EPT) - RQ1

A Pearson correlation was conducted to examine the internal consistency of the Listening part of the English Placement Test or EPT-PUiTM. The descriptive statistics and correlation results are presented in the tables below.

Table 1: Descriptive Statistics

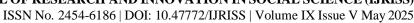| Test Paper Part	M	SD	N
Listening 1st Half Score	16.26	2.47	5423
Listening 2nd Half Score	15.08	2.73	5423

Table 2: Pearson Correlation Coefficients Between 1st and 2nd Halves of the Listening Section

Test Paper Part	Listening 2nd Half Score	Sig.	N
Listening 1st Half Score	.124	<.001	5423

The Descriptive Statistics in Table 1 above shows that the mean score for the Listening 1st Half was 16.26 (SD = 2.48), while the mean score for the Listening 2nd Half was 15.09 (SD = 2.74), with both scores based on 5423 participants. Furthermore, Pearson Correlation test results as shown in Table 2 revealed a significant positive correlation between the Listening 1st Half Scores and Listening 2nd Half Scores, $r(5423) = .124$, $p < .01$ with a nearly moderate strength of relationship. This suggests that the Listening part of the EPT still demonstrates its internal consistency, but the test setting of the reading component could have been done with some greater attention of caution in terms of ensuring that the questions are designed to ask listening questions with more homogenous levels of difficulty and its complexity in a more holistic manner.

In conclusion, the significant positive correlation between the two halves of the Listening test indicates that the test is reliably measuring the same construct across its different parts. This internal consistency is crucial for ensuring the accuracy and reliability of the overall test scores of EPT-PUiTM.

Measurement of Reading Proficiency Construct by the Reading Part of the English Placement Test (EPT) - RQ2

A Pearson correlation was conducted to examine how well the reading part of the English Placement Test (EPT-PUiTM) measures the reading proficiency construct. The descriptive statistics and correlation results are presented in the tables below.
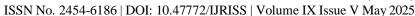
Table 3: Descriptive Statistics

Test Paper Parts	M	SD	N
Reading 1st Half Score	12.92	3.38	5423
Reading 2nd Half Score	14.00	3.40	5423

Table 4: Pearson Correlation Coefficients Between 1st and 2nd Halves of the Reading Section

Variable	Reading 2nd Half Score	Sig.	N
Reading 1st Half Score	.421	<.001	5423

The Descriptive Statistics in Table 3 above shows that the mean score for the Reading 1st Half was 12.93 (SD = 3.39), while the mean score for the Reading 2nd Half was 14.01 (SD = 3.40), with both scores based on 5423 participants. Furthermore, Pearson Correlation test results as presented in Table 4 revealed a significant positive correlation between the Reading 1st Half Scores and Reading 2nd Half Scores, $r(5423) = .421$, $p < .001$, indicating a moderate strength of relationship. This suggests that the reading part of the EPT measures the reading proficiency construct effectively.

In conclusion, the significant positive correlation between the two halves of the Reading test indicates that the test is reliably measuring the reading proficiency construct across its different parts. This internal consistency is crucial for ensuring the accuracy and reliability of the test scores of EPT-PUiTM.

Measurement of Grammar Proficiency Construct by the Grammar Part of the English Placement Test (EPT) - RQ3

A Pearson correlation was conducted to examine how well the grammar part of the English Placement Test (EPT-PUiTM) measures the grammar proficiency construct. The descriptive statistics and correlation results are presented in the tables below.

Table 5: Descriptive Statistics

Test Paper Parts	M	SD	N
Grammar 1st Half Score	7.89	1.76	5423
Grammar 2nd Half Score	6.85	2.14	5423

Table 6: Pearson Correlation Coefficients Between 1st and 2nd Halves of the Grammar Section

Test Paper Parts	Grammar 2nd Half Score	Sig.	N
Grammar 1st Half Score	.255	<.001	5423

The Descriptive Statistics in Table 5 above shows that the mean score for the Grammar 1st Half was 7.89 (SD = 1.77), while the mean score for the Grammar 2nd Half was 6.85 (SD = 2.15), with both scores based on 5423 participants. Furthermore, Pearson Correlation test results as shown in Table 6 revealed a significant positive correlation between the Grammar 1st Half Scores and Grammar 2nd Half Scores, $r(5423) = .255$, $p < .001$ with a nearly moderate strength of relationship. This suggests that the grammar part of the EPT measures the grammar proficiency construct effectively. but the test setting of the grammatical component could have been done with some additional attention of caution in terms of ensuring that the questions are designed to ask listening questions with more homogenous levels of difficulty and its complexity in its entirety.

In conclusion, the significant positive correlation between the two halves of the Grammar test indicates that the test is reliably measuring the grammar proficiency construct across its different parts. This internal consistency is crucial for ensuring the accuracy and reliability of the test scores of EPT-PUiTM.

DISCUSSION AND CONCLUSION

Internal consistency is important in measuring the test's reliability and validity as this is proven when a significant positive correlation is found between the first and second half scores in the listening, reading and grammar parts of the English Placement Tests (EPTs). The consistency functions to assure that the tests assess the proposed constructs precisely, in turn, offers reliable outcomes. This is corroborated by a study that emphasizes the significance of tailoring tests with the framework of CEFR as well as lessening cognitive load for the purpose of achieving consistency and reliability.

In accordance with the Listening section of the English Placement Test (EPT-PUiTM), internal consistency of the test is successfully established in this research. This is demonstrated when there is a significant positive correlation between the first and second half scores of the test, therefore making the questions for the listening section a valid and reliable part of the EPT-PUiTM test designed for its overall intended placement purpose. This is supported by Liu, Li and Diao (2020) who established that consistent scoring throughout tests is important in proving reliability, particularly in self-designed English listening tests. In addition, Moussa-Inaty, Sweller and Ayres (2011) further support the claim by suggesting that a reliable listening test should lessen cognitive load along with achieving sustained consistency when assessing listening tests throughout multiple sections. Besides that, Harsch and Hartig (2015) highlight the application of CEFR standard in listening tests, emphasizing the importance of internal consistency in validating test scores. Hence, the findings show that internal consistency is significant to assess the reliability of listening tests while further research may continue identifying other contributing factors in measuring reliability and test effectiveness.
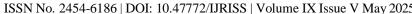
In addition, the significant positive correlation is found to extend to the Reading section of the English Placement Test (EPT-PUiTM) across the first and second half scores, thus making the questions for the reading section as valid and reliable part of the EPT-PUiTM test to achieve its designated aim. The finding aligns with the result of Baharom et al. (2021) who moot that MUET reading component acted as an important predictor of students' academic performance, implying that proficiency level of reading component can be determined through a well-structured reading test. In addition, according to Fan and Jin (2020), quality control is essential in reading placement tests, advocating that the reliability of an intended construct can be assessed through consistent correlations across various test sections. Moreover, Waluyo, Zahabi and Ruangsung (2024) illustrate that the reading proficiency level at Thai universities, which is effectively analysed using the CEFR, shows significant correlations across distinct test sections. Thus, these prove that internal consistency in reading tests is crucial to determine its reliability and validity, however, further study pertaining to additional factors could facilitate insights into test effectiveness.

Referring to the Grammar section of the EPT, which effectively analysed the grammar proficiency construct, found that there is a significant positive correlation between both Grammar first and second half scores, therefore making the questions for the grammar section as valid and reliable part of the EPT-PUiTM test to fulfil the purpose of test in assessing the students' grammatical levels of language competence. The outcome is supported by Sridhanyarat et al. (2021) who also discovered a significant internal consistency in the grammar part of STEP, an English proficiency test based on CEFR, indicating the approval of grammar test reliability that is grounded in the positive correlations shown between test halves. As reported by Hidri (2021), who analyses the application of CEFR in grammar tests, the level of grammar proficiency can be validated through consistent scoring among multiple sections of the test. The validity and reliability of a grammar proficiency construct can be evaluated based on internal consistency, thus implying its importance (Harsch & Hartig, 2015). Therefore, the findings suggest that internal consistency is considered vital particularly to establish grammar tests effectiveness, despite more investigation could be conducted to determine other influencing factors for test reliability. By ensuring the grammar test component with its established internal consistency and reliability, it helps the university's assessment administrators assess the student's language learning ability and competence at the tertiary level with greater confidence and credence (Zolkapli et al., 2025).

The findings illustrate that the different sections of EPT, namely the listening, reading and grammar sections, achieve internal consistency and effectively evaluate each respective proficiency construct. Through which, the internal consistency is deemed to be important to verify the accuracy and reliability of the EPT test scores. It is therefore concluded that all sections of the EPT-PUiTM have contributed to the overall validity and reliability of the test to be a robust tool for English placement purposes.

Having discussed the findings, the present study which investigated the English Placement Test (EPT) is believed to contribute significantly to the theoretical understanding of internal consistency in language assessment, involving different components of (i) Listening, (ii) Reading, and (iii) Grammar. The findings highlight the reliability of the test in measuring respective language proficiency constructs, in line with Liu, Li, and Diao's (2020) emphasis for test reliability and validity. Similarly, Liu et al. (2020) highlighted the importance of consistent scoring for reliable test outcomes. Other than that, the findings of the present study also support Moussa-Inaty, Sweller, and Ayres' (2011) cognitive load theory substantiating the importance of any tests to manage or reduce the cognitive processing of the test takers in a consistent and systematic manner. According to the scholars, well-constructed tests reduce cognitive load apart from maintaining measurement consistency. Not only that, the alignment of the EPT with the CEFR standards (as discussed by Harsch and Hartig (2015) further validates the effectiveness of the test in accurately interpreting language proficiency. Overall, these insights not only further exemplify the critical impact of internal consistency in language testing: they also suggest that future research should investigate other factors influencing test reliability to improve the overall effectiveness of existing language assessments.

However, one limitation of this study is that while purposive sampling enhances the relevance of the sample to the research aims, it may restrict the generalizability of the findings beyond the specific context of the participating students. Additionally, the option for diploma students to voluntarily take the English Placement Test could introduce self-selection bias, as those who opted to participate might differ in motivation and language proficiency compared to their peers who chose not to engage.

To enhance the applicability of the research, it is essential to establish a stronger connection between the study findings and practical recommendations for improving the design of the English Placement Test (EPT). The research demonstrates a successful internal consistency within the Listening section of the EPT-PUiTM, evidenced by a significant positive correlation between the first and second half scores. This not only validates the questions in this section but also reinforces their reliability for placement purposes. To capitalise on these findings, future EPT design should incorporate strategies to further reduce cognitive load during listening assessments, as suggested by Moussa-Inaty, Sweller, and Ayres (2011). This could involve simplifying task formats and ensuring that test items are contextually relevant and engaging for students.

Similarly, the Reading section exhibits strong internal consistency, aligning with findings from Baharom et al. (2021) that underscore the predictive power of reading proficiency on academic performance. To enhance the applicability of this component, it is recommended that test developers employ quality control measures that prioritise a well-structured reading framework, ensuring that each component accurately reflects students' reading abilities. This would not only improve reliability but also help identify specific areas where students may need additional support.

For the Grammar section, the study confirms the presence of internal consistency, affirming its validity as a measure of grammatical competence. To further strengthen this aspect of the EPT, it is advisable to integrate CEFR-aligned descriptors more explicitly into the test items, as highlighted by Hidri (2021). This approach can provide clearer benchmarks for proficiency levels and improve the assessment's overall effectiveness. Additionally, ongoing research should explore other factors influencing test reliability, such as the impact of instructional methods on student performance, to refine the EPT continuously. By aligning these findings with targeted recommendations, the research can effectively contribute to the development of a more reliable and valid English Placement Test, thereby enhancing its relevance and utility in educational contexts.

Future research can be done by enhancing the statistical analyses through the incorporation of qualitative feedback from both students and teachers regarding their experiences with assessments. This integration of perspectives can provide deeper insights into the quantitative data, allowing for a more nuanced understanding of how tests are perceived and experienced in real educational settings. By capturing the thoughts, feelings, and suggestions of those directly involved, researchers can better contextualize their findings, identify potential areas for improvement, and ensure that the assessments are not only valid but also meaningful and relevant to learners' needs. Overall, these recommendations are hoped to result in refined language testing practices which will then improve educational outcomes.

REFERENCES

1. Ahangari, S., & Barghi, A. (2012). Consistency of Measured Accuracy in Grammar Knowledge Tests and Writing: TOEFL PBT. Language Testing in Asia, 2. https://doi.org/10.1186/2229-0443-2-2-5.
2. Baharum, N. N., Abd Kadir, N. A., Farid, S. N. N. M., Shuhaimi, N. I. M., Rahim, W. A., Razali, A. B., & Abd Samad, A. (2021). MUET English Examination as a Predictor of Academic Achievement for TESL Teacher Trainees at a Public Teacher Education Institution in Malaysia. International Journal of Academic Research in Progressive Education and Development.
3. Bakri, H. (2022). Evaluating and Testing English Language Skills: Benchmarking the TOEFL and IELTS Tests. International Journal of English Linguistics. https://doi.org/10.5539/ijel.v12n3p99
4. Bernard, J. (1949). The Rosenzweig Picture Frustration Study: II. Interpretation. The Journal of Psychology, 28(2), 333–343. https://doi.org/10.1080/00223980.1949.9916014
5. Campbell, S., Walkem, K., Shearer, T., Walker, K., Bywaters, D., Greenwood, M., Prior, S., & Young, S. (2020). Purposive sampling: complex or simple? Research case examples. Journal of Research in Nursing, 25(8), 652–661. https://doi.org/10.1177/1744987120927206
6. Chung, S., Haider, I., & Boyd, R. (2015). The English Placement Test at the University of Illinois at Urbana-Champaign. Language Teaching, 48(3), 284-287. https://doi.org/10.1017/S0261444814000433
7. Fan, J., & Jin, Y. (2019). Standards for language assessment: demystifying university-level English placement testing in China. Asia Pacific Journal of Education, 40(3), 386–400. https://doi.org/10.1080/02188791.2019.1706445
8. Fan, J., & Jin, Y. (2020). Standards for language assessment: Demystifying university-level English

placement testing in China. Asia Pacific Journal of Education, 40(3), 386-400. https://doi.org/10.1080/02188791.2019.1706445

9. Field, A. (2013). Discovering Statistics Using IBM SPSS Statistics. Sage.

10. Fulcher, G. (1997). An English language placement test: issues in reliability and validity. Language Testing, 14(2), 113-139. https://doi.org/10.1177/026553229701400201

11. Mohamad, H. A., Abd Rashid, M. H., Abd Rahman, A. L., Zolkapli, R. B. M., Hasan, H. R., & Nath, P. R. (2024). Secondary english grade as a predictor of students' entrance english competence and programme placement. Quantum Journal of Social Sciences and Humanities, 5(5), 342-352. https://doi.org/10.55197/qjssh.v5i5.486

12. Harrington, M., & Carey, M. (2009). The on-line Yes/No test as a placement tool. System, 37, 614-626. https://doi.org/10.1016/J.SYSTEM.2009.09.006.

13. Harsch, C., & Hartig, J. (2015). What are we aligning tests to when we report test alignment to the CEFR? Language Assessment Quarterly, 12(4), 333-362. https://doi.org/10.1080/15434303.2015.1092545

14. Hidri, S. (2021). Linking the International English Language Competency Assessment suite of examinations to the Common European Framework of Reference. Language Testing in Asia, 11(1). https://doi.org/10.1186/s40468-021-00123-8

15. Kim, Y., & Kim, M. (2017). Validations of an English Placement Test for a General English Language Program at the Tertiary Level. , 20, 17-34. https://doi.org/10.20622/JLTAJOURNAL.20.0_17.

16. Krus, D. J., & Helmstadter, G. C. (1987). The Relationship between Correlational and Internal Consistency Notions of Test Reliability. Educational and Psychological Measurement, 47(4), 911–915. https://doi.org/10.1177/0013164487474006

17. Leung, C., & Jenkins, J. (2020). Mediating communication – ELF and flexible multilingualism perspectives on the Common European Framework of Reference for Languages. Australian Journal of Applied Linguistics, 3(1), 26–41. https://doi.org/10.29140/ajal.v3n1.285

18. Liu, Z., Li, T., & Diao, H. (2020). Analysis on the Reliability and Validity of Teachers' Self-designed English Listening Test. Journal of Language Teaching and Research, 11(5), 801-808. https://doi.org/10.17507/jltr.1105.16

19. Mclean, S., Kim, Y. A., Ueno, S., Raine, P., Huston, L., Nishiyama, S., & Pinchbeck, G. G. (2021). The internal consistency and accuracy of automatically scored written receptive meaning-recall data: A preliminary study. Vocabulary Learning and Instruction, 10(2), 64–81. https://doi.org/10.7820/vli.v10.2.mclean

20. Mohamad, H. A, Abd Rahman, A. L., Abd Rashid, M. H., Mohd Akhir, N., Zaraini, N. S., Rozman Azram, A. A., Hasan, H. R., Md Zolkapli, R. B. (2025). Diploma Students' Overall English Performance Based on their Different Proficiency Skills (Special Issue on Education). International Journal of Research and Innovation in Social Science (IJRISS), 9 (IIS), 1804-1816. https://dx.doi.org/10.47772/IJRISS.2025.90400136

21. Moussa-Inaty, J., Sweller, J., & Ayres, P. (2011). Improving Listening Skills in English as a Foreign Language by Reading Rather than Listening: A Cognitive Load Perspective. Applied Cognitive Psychology, 26(3), 391–402. https://doi.org/10.1002/acp.1840

22. Pallant, J. (2020). SPSS Survival Manual: A Step-by-Step Guide to Data Analysis Using IBM SPSS. Routledge. https://doi.org/10.4324/9781003117452

23. Sheerah, H., & Yadav, M. (2022). The Use of English Placement Test (EPT) in Assessing the EFL Students' Language Proficiency Level at a Saudi University. Rupkatha Journal on Interdisciplinary Studies in Humanities. https://doi.org/10.21659/rupkatha.v14n3.24

24. Sood, P. (2017). Language Proficiency and Assessing Classroom Achievement: A Literature Review. International Journal of Linguistics, 2(1), 9-16. https://doi.org/10.17161/ILI.V2I0.6933

25. Sridhanyarat, K., Pathong, S., Suranakkharin, T., & Ammaralikit, A. (2021). The Development of STEP, the CEFR-Based English Proficiency Test. English Language Teaching, 14(7), 95-106. https://doi.org/10.5539/elt.v14n7p95

26. Sufi, M. K. A., & Idrus, F. (2021). A preliminary study on localising the CEFR written production descriptor to Malaysian higher education context. Asian Journal of Research in Education and Social Sciences, 3(2), 1-15.

27. Tabachnick, B. G., & Fidell, L. S. (2019). Using Multivariate Statistics. Pearson.

https://doi.org/10.4324/9781315814919

28. Waluyo, B., Zahabi, A., & Ruangsung, L. (2024). Language Assessment at a Thai University: A CEFR-Based Test of English Proficiency Development. Reflections, 31(1), 25–47. https://doi.org/10.61508/refl.v31i1.270418

29. Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrument Penelitian Kuantitatif. Journal Tarbiyah: Journal Ilmiah Kependidikan, 7(1). https://doi.org/10.18592/tarbiyah.v7i1.2100

30. Zolkapli, R. B. M., Kenali, S. F. M., Hadi, N. F. A., Jaafar, A. J., Mohamad, H. A., Abd Rahman, A. L., & Shaharudin, N. A. D. (2025). Addressing English Grammar Learning Challenges Among Malaysian Islamic Studies Students. Malaysian Journal of Social Sciences and Humanities (MJSSH), 10(1), e003199-e003199. https://doi.org/10.47405/mjssh.v10i1.3199

31. Zolkapli, R. B. M., Kenali, S. F. M., Hadi, N. F. A., Basiron, M. K., Shaharudin, N. A. D., & Mohamad, H. A. (2024). Exploring Reasons for Learning English and Burnout Among Pre-University Students. Malaysian Journal of Social Sciences and Humanities (MJSSH), 9(1), e002670-e002670. https://doi.org/10.47405/mjssh.v9i1.2670