IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

# Impact of Question Phrasing and Cognitive Subskills Tested in EFL Assessments

[1]Golda Cherian
Assistant Lecturer, Center for Preparatory Studies
Sultan Qaboos University, Muscat, Oman

[2]Josekutty Joseph
Assistant Lecturer, Center for Preparatory Studies
Sultan Qaboos University, Muscat, Oman

***Abstract:*** Creating appropriate questions that assess different cognitive subskills of learners in reading and listening is crucial in English as a Foreign Language (EFL) education, especially in summative assessments. This study examines the differences in question difficulty levels and how the assessed subskills vary in terms of cognitive demand even when the same question format is used. A set of questions developed by the researchers demonstrates how test writers and material developers can modify question difficulty levels to differentiate learners according to their performance. The study also highlights how phrasing and cognitive demand can significantly influence fairness and reliability in classroom assessments. This theoretical study provides practical implications in the field of material creation and test development to ensure appropriateness of tasks used and fairness in the distribution of difficulty level. It further offers a conceptual model that can guide future empirical research on aligning question types with learner performance and additional research in listening and writing skills.

***Index Terms*** - question types, EFL assessment, sub skills, skimming, scanning, intensive reading, difficulty level, inference, critical thinking, test writers, material developers, learner performance, appropriate tasks, fair distribution.

## Introduction

In teaching and learning English as a Foreign Language (EFL) contexts, materials used to teach and exams used to assess students' performance play a crucial role. Different question types assess different aspects of language proficiency, measuring how well students have mastered subskills of the main language skills. Exams and classroom activities consist of various question types, including multiple-choice, true/false, short-answer, and essay questions. Nevertheless, there will be differences among these questions in assessing the different cognitive abilities of the learners from lower order thinking skills like remembering, understanding and application to higher order thinking skills like analysis, evaluation and synthesis. In EFL assessments, each question tests a different language sub-skill, like skimming, scanning and intensive reading or listening. Consequently, the difficulty level between questions can vary. Understanding item difficulty is crucial for designing effective tests. These variations in question difficulty appropriately challenge students and help educators distinguish differences in student performance. This conceptual analysis proposes a practical framework for test design, offering educators, material developers, and exam writers insights derived from the researchers' collaborative investigation.

**Different Cognitive Sub-skills in Language Learning and their Difficulty Levels**

Various cognitive sub-skills are involved in each reading and listening such as Skimming, Scanning and Intensive Reading. Skimming is a critical reading strategy to grasp the general gist or main idea of a text by rapidly reading through a text. This skill is useful when a reader needs to quickly understand the overall idea of the material. Skimming typically involves identifying keywords, phrases, and topic sentences that convey the main points of a text. Previewing a chapter or browsing through articles to identify one of interest is an example for using the skill of skimming (Grabe and Stoller, 2019). A student's ability to process essential information without being distracted by details is crucial in both academic and real-world contexts. (Brown, 2007). In simple terms, skimming means identifying what the text is generally about. Skimming is usually rated as an 'easy' skill since it involves only a quick reading of the headings, subheadings or reading the first and last sentences of the text. However, skimming can be challenging if it requires students to synthesize information through careful reading of the entire text[1].

Scanning is locating specific information swiftly that is needed to answer the questions. It is another vital reading and listening skill, to find specific information, such as names, dates, figures, or particular facts within a text. In scanning, students require to look for specific pieces of information rather than the overall idea. This skill is important when we need to deal with larger texts and specifically, during exams where swift retrieval of specific information is necessary. Scanning compels students to navigate through a text swiftly and efficiently, displaying their proficiency in collecting specific information.[2] Since it doesn't require students to read the text carefully to analyze and process information, it is rated an 'easy', lower order skill.
Intensive reading or listening involves engaging deeply with the text to analyze and process detailed information. Intensive reading and listening go beyond the surface level, requiring a much deeper level of engagement and comprehension. In intensive reading and listening, learners delve into the text, analyzing, interpreting and reflecting the information critically. Intensive reading and intensive listening help develop critical thinking, equipping students with the ability to analyze, evaluate, and synthesize information. This skill makes students draw inferences, or reading between the lines, to understand hidden meanings or ideas. It also helps them identify biases, and make logical conclusions based on the text or spoken language. This subskill assesses language learner's ability to engage with complex material and it demonstrates their higher-order thinking skills. Mastering this subskill is crucial for academic success, as it enables them critically engage with content, comprehend underlying meaning, and thereby draw informed conclusions (Macmillan English, 2023). Intensive reading or intensive listening is considered to be a difficult skill compared to skimming and scanning.

These cognitive sub-skills are not only important in academic contexts but are also essential to navigate the difficulties of daily life and professional contexts (SkillsYouNeed, 2023; Lumen Learning, 2023).
However, as Weir (2005) comments, these skills are "still theoretically constructs, with only a hypothesized existence". Therefore, it is important to make appropriate tasks for classroom materials to enhance these skills of our learners and it is more crucial to be extra careful while making questions for the tests to assess these skills. A good activity or test will have a mix of easy and difficult questions that assess all cognitive subskills of language learning. A test with a mix of question types can effectively distinguish and assess students' performance.

**Different Question Types**

Careful consideration is necessary when preparing questions in language assessments to ensure they evaluate the intended skills. Each listening and reading language test will usually have a variety of question types, namely, multiple-choice questions, true/false/not given questions, one-word answer questions, short answer questions, fill in the blank questions, summary completion and so on. According to Brown & Hudson (1998), multiple-choice and true/false questions can effectively test vocabulary, grammar, and basic comprehension, permitting clear and objective grading while short answer tasks are suitable to assess learners' ability to produce language and express themselves concisely and coherently. Similarly, the Center for Teaching Innovation at Cornell University advocates for the use of different question types because of the fact that different students may perform differently with different formats. The center suggests incorporating a mix of

[1] *While generally considered easy, advanced skimming techniques require evaluating multiple sources for synthesis and comparison.*
[2] *Scanning is often confused with skimming; however, scanning focuses on retrieving specific details rather than overall understanding.*

multiple-choice, short-answer, and essay question types to ensure a more comprehensive and robust evaluation of student learning.

**The Right Task for EFL Activities and Assessments**

Brown & Hudson (1998), writes that the selection of task types, whether it should be a scanning question or an inference question depends mainly on the assessment objectives and the specific language skills being measured. We can never argue that all multiple-choice questions or true/false questions are the best for checking the scanning. Multiple-choice questions (MCQs) can assess a wide range of cognitive skills, depending on how they are structured.[3] Likewise, summary completion does not always promote critical thinking skills. It can sometimes be a scanning or skimming question depends on how the questions are framed. It is crucial to have a balanced question paper with a mix of different task types to assess the different language skills and sub-skills. By administering carefully written tests, we can ensure fair and effective assessments. The British Council website reminds educators that while certain question types are versatile and can test various skills, they may not be very effective for testing specific items like grammar or vocabulary. Therefore, aligning exam questions with students' language level and selecting question types that match the language proficiency and learning objectives of students is essential to ensure accuracy and fairness. Test writers should always make sure that the questions are aligned with students' level of language proficiency and the questions are of mixed difficulty levels with varied question types that assess different language subskills.

**Literature Review**

Xiong et al. (2022) in their study, provided an overview of current approaches in estimating difficulty level of questions from textual content and they conducted a survey on these approaches. The findings of the survey highlight the importance that it is crucial to accurately gauge question difficulty to create assessments that are appropriately challenging to students. Using questions with varied difficulty levels is necessary to differentiate students' performance at varying levels of achievement.

The study conducted by Zhao and Lee (2022) investigated the impact of teaching listening subskills, especially micro- and macro-skills, using a speaking-listening model within a computer-mediated communication environment. 112 Chinese tertiary students with intermediate English proficiency participated in the study. The researchers found that that practicing listening subskills orally in advance to doing listening activities improves students' performance.

Eren and Mocan (2019) examined how the ordering and difficulty level of questions affect student performance. The study revealed that placing questions with higher difficulty level at the beginning of a test could cause higher dropout rates and lower results. The study sheds light into the importance of sequencing question strategically to retain student engagement throughout the test and it will also help carry out accurate assessment.

Guan (2019) examined the development of listening proficiency in understanding various spoken genres among the EFL learners in Hong Kong's tertiary education. The study identified underlying subskills in the Diagnostic English Language Tracking Assessment (DELTA) which was the test used to assess students' proficiency. The study examined the difficulty levels of these subskills and the findings provide useful insights into how an effective listening test can be designed.

Karakoc (2019) investigated the alignment between reading and listening subskills and their application in EFL books and proficiency tests. The study found ten common subskills shared between reading and listening. The subskills are also exclusive to each. The findings of this study help educators develop both instructional materials and balanced tests that accurately measure students' proficiency in the subskills.

A study conducted by Couch et al. (2018) compared Multiple-True-False (MTF) and Free-Response (FR) question types to analyse how different question types can affect students' understanding. The findings revealed that MTF questions could bring forth mixed conceptions, whereas FR questions highlighted partial

[3] *For example, a factual MCQ requires recognition, while a well-designed MCQ can assess inferencing or problem-solving.*

understandings. This study indicates the importance of including diverse question formats in assessing student comprehension.

Alfonseca et al. (2014) explored the discrepancies between teachers' estimations and students' perceptions of question difficulty. The findings of the study indicated that students often viewed questions as more complex than teachers estimated. It highlights the need for educators, content developers and test writers to carefully look into question phrasing and difficulty level to match with the proficiency levels of students.

Aryadoust et al. (2012) conducted research on whether listening subskills could be distinguished within assessments. The findings of the study indicated that listening subskills in the test were identifiable and divisible. It supported the idea that specific subskills in listening can be separately assessed. This study has major implications for the test writers and material developers showing that tests can be prepared to assess target subskills of listening and likewise, reading effectively.

Roediger and Butler (2011) investigated the benefits of employing different question formats, like cued recall, multiple-choice, and true/false question patterns, to improve learning and assessment. The study found that diverse question types can cater to various learning styles and support a much more comprehensive assessment of student learning.

Johnson (2009) in his article advises that right selection and careful phrasing of questions while creating the tests will allow educators to adjust the difficulty level. Having questions with different difficulty level is crucial for differentiating student performance. The author highlights the relevance of using distractors and changing question complexity to create various difficulty levels, to enable differentiation in all assessments.

To summarize, to differentiate student performance and ensure validity and fairness, a well-designed assessment or a classroom activity should include a variety of question types that have carefully adjusted difficulty levels. Aligning theoretical frameworks with practical applications in test design enhances the evaluation of language competencies. A well-balanced approach to teaching and testing will cater to the needs of diverse learning styles and provide a comprehensive as well as an equitable evaluation of student achievement.

**Methodology**

This study employs a qualitative analytical approach to examine how question phrasing influences difficulty levels in EFL assessments. Two sets of reading comprehension questions were created for the same text, one with lower difficulty level (Set 1) and the other with higher difficulty level (Set 2), to demonstrate how changes in phrasing and cognitive requirement alter difficulty and the underlying reading subskill and consequently, influence learner performance.

The researchers designed ten questions for each text, ensuring that the same passage was used while altering the phrasing and cognitive focus of the items. The first set consisted of questions that tested lower-order skills such as skimming and scanning, while the second set included higher-order questions that required inference, analysis, and evaluation.

**Criteria for Classifying Questions into Easy vs. Difficult**

Four main parameters guided the preparation and qualitative analysis of the questions: (1) Explicitness of information in the text, (2) Cognitive processing required, (3) Plausibility of distractors, and (4) Type and intent of the question. The researchers analysed each question according to these parameters to determine its difficulty level and the subskill it primarily assessed. The classification of questions was based on the following factors given below:

1. **Explicitness vs. Implicitness:** Easy questions had explicit answers directly stated in the text, while difficult questions required inference, synthesis, or paraphrased understanding.
2. **Cognitive Processing:**
   - Set 1 (Easy) targeted easier skills, such as skimming, scanning, and fact identification which requires only a quick reading of the passage and locating the answer in the text.
   - Set 2 (Difficult) assessed difficult skills, including critical thinking skills like analysis, evaluation, and inference which requires intensive or careful reading of the passage. Answers are not directly retrievable from the text.
3. **Complexity of Distractors:**
   - Set 1 used clear, straightforward distractors, making elimination easier.
   - Set 2 included plausible but misleading distractors requiring deeper comprehension.
4. **Question Type Adaptation**: While the format (MCQ, T/F/NG, Short Answer) remained consistent, the phrasing was modified to increase cognitive demand in Set 2.

This classification framework was applied to analyze the impact of question rewording on difficulty levels and identify best practices for test design.

The reading text and the two sets of questions that differ in difficulty level created by the researchers are given below:

***Table 1: Reading Text***

**Reading Text:**

Sleep deprivation is a growing problem that affects millions of people worldwide. Some health experts define sleep deprivation as getting fewer than seven hours of sleep each night, but others believe the ideal number depends on age, lifestyle, and even genetics. Some studies suggest that teenagers need up to ten hours for healthy growth. Insufficient sleep disrupts the body's normal functioning.

One major consequence of poor sleep is the decline in mental performance. People who sleep less tend to lose focus, react slowly, and forget simple information. They may also make poor decisions and have difficulty solving problems. Sleep deprivation weakens concentration, memory, and reasoning ability, which are essential for daily tasks and learning.

Lack of sleep also harms physical health. When people sleep less, their bodies produce stress hormones, which can increase blood pressure and raise the risk of heart disease. Some research link chronic sleep loss to obesity and diabetes and have also found that sleep-deprived individuals often crave sugary or high-fat foods, which can further damage their health over time. Sleep deprivation increases the risk of catching infections by 40 percent, though that statistic may depend on lifestyle and nutrition. Human body cannot repair tissues or maintain a strong immune system without adequate rest.

The emotional and social effects of sleep deprivation are equally serious. People sleep less often feel irritable, anxious, or even depressed. They often argue with others, which affects relationships at home and work. A recent report claimed that accidents increase by 10 percent after night shifts. In short, poor sleep reduces emotional control, damages communication, and lowers overall well-being.

*Table 2: Two Sets of Reading Test with Different Difficulty Levels*

**Set 1 (Easy Questions)**

**Choose the correct answer:**

1. What is the main idea of the passage?
   A. Sleep deprivation affects only mental performance.
   B. Lack of sleep can harm a person's mind, body, and emotions.
   C. Getting fewer than seven hours of sleep is always dangerous.
2. According to the passage, what happens when people sleep less?
   A. They become alert.
   B. They lose focus.
   C. Their memory improves.
3. What can happen to people's physical health because of sleep deprivation?
   A. It can raise blood pressure and cause heart disease.
   B. It can make them immune to infections.
   C. It helps them maintain a healthy weight.
4. What emotional effect of sleep deprivation is mentioned in the passage?
   A. Hatred and revenge
   B. Anxiety and irritability
   C. Anger and jealousy

Write **T (True)** if the statement agrees with the passage, **F (False)** if it contradicts the passage, or **NG (Not Given)** if there is no information:

5. Teenagers need the same amount of sleep as adults. _____
6. Sleep deprivation can increase the risk of catching infections. _____
7. People who sleep less eat more vegetables. _____
8. Accidents decrease after night shifts. _____

**Short Answer (Write No more than THREE words):**

9. What hormone does the body produce more of when sleep is reduced? _____
10. Which body system cannot be maintained without adequate rest? _____

**Set 2 (Difficult Questions)**

**Choose the correct answer:**

1. Which statement best summarizes the passage?
   A. Sleep deprivation negatively affects mental, physical, and emotional health.
   B. Sleep deprivation improves emotional awareness but weakens memory.
   C. Emotional well-being is the only area influenced by lack of sleep.
2. Which scenario best shows the cognitive effect of sleep deprivation?
   A. A person eating high-fat foods late at night
   B. A worker arguing with colleagues after a shift
   C. A student forgetting simple facts during an exam
3. Which statement best explains how sleep deprivation affects social relationships?
   A. It encourages individuals to isolate themselves for better focus.
   B. It causes irritability and poor emotional control, leading to conflicts.
   C. It reduces empathy because people spend less time on social media
4. If night shifts are increased, what is the most likely outcome according to the text?
   A. Accidents and fatigue may increase, reducing work efficiency.
   B. Workers will adapt and develop stronger concentration skills.
   C. Employees may become physically healthier but socially withdrawn.

**Write T (True), F (False) or NG (Not Given):**

5. Genetic factors may influence how much sleep a person needs. _____
6. People who sleep less are guaranteed to develop heart disease. _____
7. The 40 percent increase in infection risk applies to every individual. _____
8. Emotional and physical effects of poor sleep are temporary. _____

**Short Answer (Write No more than TWO words):**

9. Which hormone is linked to sleep deprivation and increased blood pressure? _____
10. What main organ system fails to repair properly without adequate rest? _____

**Analysis:**

Each question in both sets of the Reading Test were analyzed qualitatively for difficulty level based on the subskill it is meant to be tested. Tables were prepared to present question-wise analysis of the task type, cognitive level, and justification for each item. The results of the analysis are given below.

***Table 3: Question-wise Analysis for Question Types, Task Types and Difficulty level- Set 1***

| Question No | Question Type | Task Type (Subskill) | Difficulty Level | Interpretation/ Justification |
|---|---|---|---|---|
| 1 | Multiple Choice | Skimming | Easy | This item requires identifying the main idea of the passage by skimming for overall meaning rather than specific details. It tests global comprehension at a lower cognitive level. |
| 2 | Multiple Choice | Scanning | Easy | The question focuses on locating factual information ("what happens when people sleep less") directly stated in the text. Students can easily scan for explicit detail ("lose focus"). Also, distractors are very obvious. |
| 3 | Multiple Choice | Scanning | Easy | This item checks comprehension of a clearly mentioned physical consequence ("raise blood pressure"). The answer is found verbatim in one sentence, making it a simple retrieval task. Also, distractors are very obvious. |
| 4 | Multiple Choice | Scanning | Easy | The question targets an emotional effect stated explicitly in the passage ("irritable, anxious"). Learners only need to recognize and match information, requiring minimal reasoning. |
| 5 | True/False /Not Given | Scanning | Easy | Students verify factual accuracy ("Teenagers need the same amount of sleep as adults") by locating and comparing statements. This involves literal understanding without inference. |
| 6 | True/False /Not Given | Scanning | Easy | This question checks understanding of a clear fact ("Sleep deprivation can increase infection risk"). As the idea is directly stated, the task remains at the recognition level. |
| 7 | True/False /Not Given | Scanning | Easy | The statement is easily identified as false through direct scanning ("crave sugary or high-fat foods"). It requires minimal interpretation or paraphrasing. |
| 8 | True/False /Not Given | Scanning | Easy | Learners confirm information ("accidents increase by 10 percent") stated explicitly. The question is factual, testing surface-level comprehension. |
| 9 | Short Answer | Scanning | Easy | Students can directly lift the phrase "stress hormones" from the passage. The task measures the |

| | | | | |
|---|---|---|---|---|
| | | | | ability to locate precise information (literal comprehension). |
| 10 | Short Answer | Scanning | Easy | The item asks for a specific phrase ("immune system") clearly mentioned in the text. It tests recognition and word-level recall rather than reasoning. |

***Table 4: Question-wise Analysis for Question Types, Task Types and Difficulty level- Set 2***

| **Question No** | **Question Type** | **Task Type (Subskill)** | **Difficulty Level** | **Interpretation/ Justification** |
|---|---|---|---|---|
| 1 | Multiple Choice | Skimming | Difficult | This item requires integrating multiple sections of the text (mental, physical, and emotional effects) to identify the best summary. It assesses analytical reasoning and synthesis, beyond simple recall. |
| 2 | Multiple Choice | Intensive Reading | Difficult | Students must interpret a situation that represents a cognitive effect ("forgetting simple facts"), linking a scenario to textual meaning. This involves applying knowledge to a new context. |
| 3 | Multiple Choice | Intensive Reading | Difficult | Learners infer how emotional and social factors interact ("poor emotional control – leads to-conflicts"). The answer is not directly stated but must be deduced from connected ideas, raising complexity. |
| 4 | Multiple Choice | Intensive Reading | Difficult | This question requires applying textual information ("accidents increase after night shifts") to predict an outcome. It involves evaluating consequences, demonstrating higher-order reasoning. |
| 5 | True/False /Not Given | Intensive Reading | Difficult | Students infer meaning from "depends on age, lifestyle, and genetics" to determine if genetic factors influence sleep. It requires comprehension beyond literal recall. |
| 6 | True/False /Not Given | Intensive Reading | Difficult | The question demands analysis of the phrase "can increase blood pressure and risk of heart disease" to evaluate the truth value. Students must notice modal language ("can") to avoid misinterpretation. |
| 7 | True/False /Not Given | Intensive Reading | Difficult | The item tests the ability to evaluate the scope of a statistic ("though that may depend on lifestyle"), requiring interpretation of conditional information. |
| 8 | True/False /Not Given | Intensive Reading | Difficult | Learners must infer that emotional and physical effects are long-term, not temporary, demonstrating inferential comprehension and textual reasoning. |
| 9 | Short Answer | Intensive Reading | Difficult | This question links cause and effect between "sleep deprivation," "stress hormones," and "blood pressure." Students synthesize information from two sentences to answer accurately. |
| 10 | Short Answer | Intensive Reading | Difficult | Learners must interpret the physiological effect of poor sleep ("immune system fails to repair"). It requires comprehension of body processes and applying conceptual understanding from the passage. |

**Findings of the Study**

The comparison of the two sets of reading tests and the qualitative analysis of each question revealed distinct patterns in difficulty and subskill distribution. The researchers observed that question phrasing significantly affects the level of cognitive processing required and, consequently, the perceived difficulty of the item. Comparison of the two reading tests and analysis of each question reveal the following key findings:

1. Reading Test 1 is an easy test since students can easily locate the answers and answers are explicitly given in the text.
2. Reading Test 2 is a difficult test since students need to use their skill for intensive reading to think critically to process, analyze and infer information to answer the questions correctly.
3. It is not the passage that makes the test easy or difficult, rather the questions can make a test easy or complex. Same passage can be used for creating an easy or a difficult test.
4. It is evident from the analysis that phrasing of question can make a question easy or difficult.
5. The question type doesn't affect the cognitive processing or the subskill it assesses, but the difficulty level of the question decides the thinking process involved in answering a question. In other words, a Multiple-Choice Question can be easy or difficult and likewise, it can evaluate a student's ability to scan or infer information from a passage.
6. When the phrasing of a question changes, the difficulty level and thereby the subskills it assesses also change.
7. It is evident that misalignment in difficulty level can lead to unfair and poor assessment. A test consisting solely of easy or difficult questions will fail to differentiate student performance.

To summarize, 'Set 1' questions primarily assessed lower-order subskills such as locating information or recognising explicit meaning while 'Set 2' questions involved deeper cognitive operations like inference, evaluation, and synthesis. The researchers noted that even when the question format remained identical, subtle changes in wording altered the skill and effort required for comprehension.

This comparison demonstrates that difficulty level is not determined by the question type itself (e.g., multiple choice, true/false, short answer), but rather by the cognitive load created through phrasing and information placement. The analysis confirms that fair and valid assessment design depends on a balanced distribution of question difficulty and subskill representation.

**Conclusion**

This study highlights the role of question phrasing in determining assessment difficulty. The comparison of the two reading sets demonstrates that question phrasing has a significant impact on the cognitive subskills being tested. Even when identical question types are used, differences in wording can shift the cognitive demand from lower-order to higher-order thinking. The researchers found that this shift influences not only the perceived difficulty level of a question but also the specific reading processes that learners employ.

As demonstrated, the findings of the study confirm that the same test format can measure entirely different skills depending on the phrasing of the question. The same passage can generate tests of varying difficulty levels, which reinforces that test design, rather than content alone, determines the difficulty level of a test. Therefore, test developers must strategically phrase and structure questions to align with learners' proficiency levels and promote fair assessment practices. By modifying question phrasing, test writers can control cognitive demand, ensuring that assessments are neither too simplistic nor excessively complex. A well-balanced test includes a mix of easy and challenging questions to effectively measure student comprehension and proficiency.

This study supports the view that validity in language assessment depends not solely on test format, but on the alignment between the task's linguistic features and its cognitive requirements. In addition, the researchers emphasize that a balanced mix of easy and difficult items enhances fairness and reliability, ensuring that assessments differentiate learners accurately across proficiency bands. For example, a multiple-choice question can assess either simple recall or deeper inferential reasoning based on how it is worded. The researchers highlight that item construction must therefore be guided by clear objectives related to subskill focus and learner proficiency level.

**Implications for Test Design & Material Development**

1. While designing teaching materials and creating tests for assessments, it is crucial to ensure a variety of tasks are included to cater to different cognitive subskills of language learning.
2. A fair distribution of questions with balanced difficulty levels is necessary to differentiate students' achievement.
3. Tests and materials should be suitable for students' language proficiency. Higher-order, difficult questions requiring intensive reading and listening are not suitable for beginners or students with lower proficiency levels.
4. It is important to be aware that how the phrasing of questions can affect the difficulty level and the skill it is meant to be assessed.

Creating appropriate materials and tests in EFL contexts is crucial and hence, test writers and material developers are obliged to ensure fairness and balance in the difficulty level of their creations. Adjusting question difficulty to match student proficiency is vital in material and test creation to ensure accurate, and fair language learning and assessment.

**Scope for Future Research**

1. Future studies could empirically validate these findings with actual student performance data to quantify how question phrasing influences comprehension accuracy.
2. This study is limited to Reading skill, it can be extended to Listening and Writing Skills.

To conclude, the results of this conceptual analysis reinforce the importance of careful test design in English as a Foreign Language (EFL) education. By modifying question phrasing, test developers and material designers can control the level of cognitive engagement required and thus produce more equitable assessments. The researchers recommend that future studies extend this analysis to listening and writing tasks to examine how phrasing influences the assessment of productive and receptive skills alike, and to verify these findings through empirical research.

**References**


1. Alfonseca, E., Pérez, D., Rodríguez, P., & Yanes, J. (2014). Automatic classification of question difficulty level: Teachers' estimation vs. students' perception. *Proceedings of the 2014 International Conference on Educational Data Mining*, 11–18.
2. Aryadoust, V., Goh, C. C. M., & Kim, L. O. (2012). An investigation of differential item functioning in the MELAB listening test. Language Assessment Quarterly, 9(3), 267–291. https://doi.org/10.1080/15434303.2011.636525
3. British Council. (n.d.). Test question types in language assessment. *Teaching English*. Retrieved from https://www.teachingenglish.org.uk/professional-development/teachers/assessing-learning/articles/test-question-types
4. Brown, H. D. (2007). *Principles of language learning and teaching* (5th ed.). Pearson Education.
5. Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. *TESOL Quarterly, 32*(4), 653–675.
6. Center for Teaching Innovation. (n.d.). Asking good test questions. *Cornell University*. Retrieved from https://teaching.cornell.edu/teaching-resources/assessment-evaluation/asking-good-test-questions
7. Couch, B. A., Hubbard, J. K., & Brassil, C. E. (2018). How question types reveal student thinking: An experimental comparison of multiple-true-false and free-response formats. *CBE—Life Sciences Education, 17*(2), ar26.
8. Edutopia. (n.d.). How to create better tests that differentiate. Retrieved from https://www.edutopia.org/better-tests-differentiate
9. Eren, O., & Mocan, N. (2019). Understanding performance in test taking: The role of question order and difficulty. *EdWorkingPapers*.
10. Grabe, W., & Stoller, F. L. (2019). *Teaching and researching reading* (3rd ed.). Routledge.
11. Guan, Y. (2019). *An assessment-based study of listening sub-skills using Rasch measurement* [Doctoral dissertation, James Cook University]. ResearchOnline@JCU. Retrieved from https://researchonline.jcu.edu.au/63730/
12. Karakoc, D. (2019). Reading and listening comprehension subskills: The match between theory, coursebooks, and language proficiency tests. *International Journal of Language Academy, 7*(4), 382–399. Retrieved from https://files.eric.ed.gov/fulltext/EJ1239087.pdf

13. Kim, Y., & Crossley, S. A. (2019). Examining the notion of listening subskill divisibility and its implications for second language listening. *Language Testing, 36*(1), 57–83. Retrieved from https://www.researchgate.net/publication/277960581
14. Lumen Learning. (2023). Introduction: Critical thinking, reading, & writing. Retrieved from https://courses.lumenlearning.com/college-readiness/chapter/introduction-critical-thinking-reading-writing/
15. Macmillan English. (2023). Encouraging extensive reading to advance intensive reading. Retrieved from https://www.macmillanenglish.com/encouraging-extensive-reading-to-advance-intensive-reading
16. Office of Educational Assessment. (n.d.). Item analysis: Understanding difficulty levels in test questions. *University of Washington*. Retrieved from https://www.washington.edu/assessment/scanning-scoring/scoring/reports/item-analysis/
17. Pishghadam, R., & Tabatabaeyan, M. S. (2021). Enhancing learning potential score in EFL listening comprehension through dynamic assessment. *Language Testing in Asia, 11*(1), 1–22. Retrieved from https://languagetestingasia.springeropen.com/articles/10.1186/s40468-021-00126-5
18. Roediger, H. L., & Butler, A. C. (2011). The value of using tests as learning events. *Educational Psychology Review, 23*(3), 399–411.
19. SkillsYouNeed. (2023). Critical reading and reading strategies. Retrieved from https://www.skillsyouneed.com/learn/critical-reading.html
20. Weir, C. J. (2005). *Language testing and validation: An evidence-based approach*. Palgrave Macmillan.
21. Xiong, Y., Zhao, S., & Van der Weide, T. (2022). A survey on recent approaches to question difficulty estimation. *ACM Computing Surveys (CSUR), 55*(5), 1–37.
22. Zhao, C., & Lee, H. (2022). Teaching EFL learners listening subskills with a speaking-listening model in a computer-mediated communication setting. *Frontiers in Psychology, 13*, 836013. Retrieved from https://www.frontiersin.org/articles/10.3389/fpsyg.2022.836013/full