

Beyond the Bubble: Unveiling the Multifaceted Landscape of Test Wiseness and Their Operationalization Among English-Language Majors

Abdullah A. Al Fraidan

Department of English Language, College of Arts, King Faisal University, Al Ahsa, Saudi Arabia

Abstract—Test wiseness is viewed as learner tactics applied in a test situation without processing the test construct. Existing studies often on test-taking strategies attribute a single function to them, despite some acknowledging their multifunctionality. This study aims to provide a new reflection of test-taking strategies and test wiseness based on their multifunctionality and how they are operationalized. Forty male English-language majors at King Faisal University participated in a longitudinal study over two years. Think-aloud protocols and retrospective interviews were used for the qualitative analysis. The results showed that students used different test-taking strategies in chains and clusters, with various functions (cognitive, metacognitive, affective, and social). The study also analyzed these strategies in terms of how they were operationalized. The study concludes that these strategies may work at cross-purposes with the goal if they are operationalized poorly, or if the choice of strategies is inappropriate for the task. In addition to theoretical and pedagogical implications for test-taking strategies, this study suggests recommendations for future studies.

Index Terms—test-wiseness, test-unwiseness, test-deviuousness, EFL learners' strategies

I. INTRODUCTION

Language learning strategies (LLS) is still a constant field with consistent findings worldwide (Alyami, 2011). However, research on Test-taking strategies (TTS) has not produced new reflections that address the limitations of previous studies, such as the persistent use of the same research tools and uncovering the reality of those LLS (Cohen et al., 2023a). Cohen et al. (2023a) called for new reflections in LLS, which covered TTS as a subject within LLS. Most research on TTS has been questionnaire-based, and only looked at LLS on a macro level without investigating the complexities of using questionnaires to generate claims regarding LLS or TTS. The complexity of how learners look at LLS and TTS was never investigated, within the framework of whether they were validating what they report was their actual TTS or only their own beliefs about those strategies.

Previous studies classified TTS under categories and sub-categories (i.e., test management strategies, strategies for reaching an answer) (Cohen, 2011), signaling that a single strategy has only one function. While this prevailing notion persisted in the literature for several decades, it has become apparent that a strategy can have more than one function (Cohen et al., 2023a). Thus, this study seeks to contribute novel insights by delving into the self-reports of Test-Taking Strategies (TTS) among Saudi English as a Foreign Language (EFL) learners. Specifically, it aims to explore how these learners employ diverse multifunctional TTS, often operating in interconnected chains and clusters (Al Fraidan, 2019), and their potential influence on test scores. Notably, this research marks the first attempt, to the best of the researcher's knowledge, to analyze the strategic operationalization of multifunctional TTS, distinguishing between wise and unwise application in the pursuit of reaching answers (Al Fraidan, 2014; Al Fraidan & AlSalman, 2023).

II. LITERATURE REVIEW

Definition of TTS

TTS is defined as a cognitive process that facilitates or solves a test item/task (Al Fraidan, 2014). However, several TTS are neutral, and not inherently facilitating or debilitating. These positive or negative outcomes depend on how they are operationalized.

According to Cohen (1994, p. 119), TTS involves strategies that learners implement while answering test tasks, "viewed simply as learner strategies applied to the area of assessment." Allan and MacLennan (1997) present the term "test-taking strategy" as the capability to mark and explore the weaknesses in the objectives of the language test items, to attain keyed answers. This does not use either the knowledge or abilities of the things mentioned that were to be measured in in the test. TTS allows test-takers to utilize the design and features of a test to escalate their scores in test-taking conditions. For instance, they can comprise on reading the test's guidelines, distribute the time on each section

wisely, employ some of the keywords in the questions, postpone tackling challenging items, and revise the entire task to ensure the correctness of the answers (Rogers & Harley, 1999). This process is referred to as the test-wise strategy.

TTS and construct validity

TTS has been viewed as either construct-related or construct-unrelated strategies (Al Fraidan, 2014). As Cohen et al. (2023b) stated, TTS is either a language-use strategy or test deviousness (a newly introduced term for test wiseness). The first is when test takers use rehearsal, retrieval, coverage, and communication strategies. This implies that the individual performed actions related to its construct during the test task. The latter occurs when a test-taker does not refer to his/her essential knowledge in the process of taking a test, but rather avoids performing the construct to reach an answer (Cohen et al., 2023b). The term test deviousness is introduced later to clarify the ambiguity of the term test-wiseness, and suggests positive results and student cleverness. However, the word “avoid” in the later definition is not accurate. While it is true that students avoid showing their true knowledge in some cases, in others, test takers do not usually intentionally do so. Therefore, this definition could be considered inaccurate and misleading. In one case, a test taker chooses the correct answer and then changes it into an incorrect answer because they were conflicted between the two answers. Therefore, they are neither hiding their true knowledge of the tested construct nor avoiding showing it. This study proposes a new definition of TTS that includes test (un)wiseness.

Test-unwise strategies occur when students apply construct relevant strategies (Al Fraidan, 2011, 2014). This mostly occurs during language tests, when students have the knowledge and ability to answer the questions correctly. However, in the final stage of the answering process, students attempt to incorrectly change some correct answers (Al Fraidan, 2014; Al Fraidan, 2019; Al Fraidan & AlSalman, 2023). Thus, test-unwise strategies have a crucial role as real barriers to language test validity (Yang et al., 2019; Kivity & Huppert, 2016).

In summary, a comprehensive understanding of both test wiseness and test unwiseness contributes to refining TTS definition and coaching by highlighting effective approaches and addressing potential pitfalls, ultimately fostering improved test performance and outcomes.

Classifications of TTS

Previous studies have relied on the known classifications of TTS: (a) test management strategies, (b) test learner strategies, and (c) test-wiseness strategies (Cohen & Upton, 2006; Cohen, 2011). This classification has been prevalent for a long time, with few modifications, like introducing test unwiseness as a fourth category (Al Fraidan, 2014). However, it is not exclusive or illustrative of the reality behind TTS.

One persistent problem with this classification is the overlap of three or four categories. For example, using semantic knowledge to exclude distractors in a multiple-choice vocabulary test can be seen as a TTS for saving time (test management) and a wise strategy to reach an answer.

One prevailing categorization of TTS is given by Phakiti (2003), who introduced two main categories—cognitive and metacognitive strategies—to explain the discrepancy in the performance of 348 EFL Thai learners. With Cognitive strategies include deductive reasoning and summarizing to comprehend the intended language (Oxford, 1990). Meanwhile, metacognitive strategies play a role in monitoring cognitive strategies (Flavell, 1981). This means revising, testing, and examining any steps that need to be solved, assessing the strategies used by the learner themselves, preparing for further movements, and observing the use of actions (Baker & Brown, 1984). Limiting TTS to only two main categories is inaccurate as this would neglect other strategies, such as affective (e.g., overcoming test anxiety) and social strategies (e.g., asking for help from the teacher). Although this work holds significance in the realm of TTS, the existing classification proves inadequate, failing to fully encapsulate the reality and nuances of TTS usage.

Other studies worldwide have classified strategies according to skill (AlSafi, 2019; Rupp & Ferne, 2006). This kind of categorization has been ongoing for decades, and has been used by different researchers (see Addamegh, 2006; Alyami, 2011; Khan & Rasheed, 2019; Killi & Leu, 2019). The majority of previous TTS studies have predominantly employed questionnaires for classifying TTS into various categories and themes. However, a limited number of studies utilized think-aloud protocols, proving instrumental in capturing a deeper understanding of how TTS relate to construct and score validity through previously unexplored areas like multifunctionality, clustering, wiseness of operationalization, which this study is trying to substantiate.

Learners' utilization of TTS varies according to the numerous elements by which the nature of the given task needs to be fulfilled. Hence, it is essential to understand the mandatory test-taking strategies for EFL students in various types of tests such as speaking, reading, listening, and writing (AlSafi, 2019).

Building upon the aforementioned definitions and classifications of TTS, it becomes apparent that weaknesses exist, particularly in the predominantly questionnaire-based approaches. Conversely, verbal-protocol studies, though valuable, have their own limitations, often neglecting significant aspects of TTS, such as multifunctionality. Thus, a more nuanced and comprehensive approach is needed to address these gaps and offer a thorough understanding of the multifaceted nature of TTS.

TTS and the PPP model

One approach used to conceptualize TTS is the presentation, process, and production (PPP) model (Figure 1) (Al Fraidan, 2011). Teachers attempt to present the test task as providing students with a chance to practice what has been learned through controlled activities, while the learners apply what has been taught (process) to come up with answers to complete the task (product) (Flippo et al., 2018). However, this model is one method in which the second and last

steps are affected by the first step (presentation). Al Fraidan (2012) mentioned that there might be a two-way interaction between the product and process stages. At the product stage, learners may retrospectively modify, confirm, or change the product because of the different processes that occur before a final answer has been chosen. Thus, the answers chosen by the students involves operations such as crosschecking, changing, and reviewing solutions that do not fit, the learner's perspective. Therefore, through such a process, the dual-way interaction explains the unwise operationalization of strategies, as seen in vocabulary tests with multiple-choice questions, where students change the correct answer for the wrong ones (Al Fraidan & AlSalman, 2023).

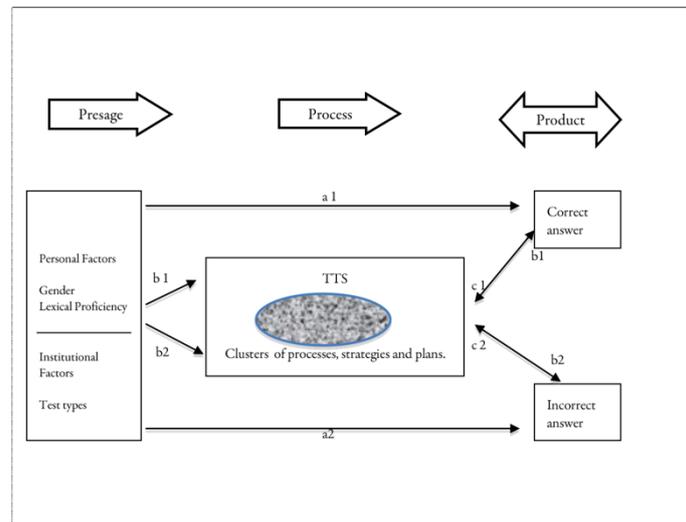


Figure 1. Al Fraidan's (2011) PPP Model

The PPP model serves as an indicator illuminating how TTS cluster and exhibit multifunctionality. The interplay among its three stages offers insights into the ways TTS are operationalized, either wisely or unwisely.

Current insights on TTS

There has been linear progression in the field of LLS in general, and TTS in particular, for several decades. TTS has been examined from similar perspectives for several decades. Nourdad and Ajideh (2019) noted that they are accurate tools for test-takers to obtain better test results by using cognitive and metacognitive strategies, where a constructive bond is strongly tied between test performance and TTS. Others see these strategies as learning or learner strategies, which are only used in test situations (Cohen, 2013; Al Fraidan, 2019). They have also been classified in different ways according to their nature (cognitive, metacognitive), skill (reading, writing, vocabulary strategies), task (TTS), or strategy (strategies for handling the gap in a gap-filling test). Even very recent studies (Han, 2023) focus on previous arguments about how the test task and student level affect TTS usage.

High-level students experienced metacognitive success, whereas low-level students experienced cognitive failure. However, highly successful test-takers are prone to using the metacognitive strategies compared to moderately successful students, who, in turn, implement the set-aside strategy compared to the unsuccessful test-takers. The PPP model shows how students can revise their answers through two-way engagement. Hence, it is essential to encourage students to use the right procedures when taking the test.

The focus should now be on the details of other characteristics of TTS that were established in earlier studies. Some of these characteristics are strategy clustering, being multifunctional, and its wise or unwise operation, which the current study attempts to explain thoroughly.

III. METHODOLOGY

Many strategy elicitation studies have mainly relied on questionnaires, which are only one of the nine methodologies implemented when evaluating learning strategies (Oxford, 2010). This method includes real task verbal reports, individual interviews, observation, narratives such as students' diaries, color coding, discussion, group interviews, discourse analysis, and students' portfolios. Most studies have used questionnaires based on Schmitt (1997) and Oxford's (1990) taxonomies. Think-aloud protocols with careful, valid, and reliable coding and analysis are rarely used in TTS studies, because they are difficult to implement.

Think-aloud verbal reports are a vital method for investigating TTS. Many researchers have asserted that think-aloud is the best option so far to reach the most accurate results regarding the usage of TTS. From my own perspective, the need for additional examination of TTS through think-aloud protocols remains because they are not widely used in TTS literature.

Participants

The researcher aimed to collect as much data as possible to conduct a valid analysis of the claimed characteristics of TTS. 40 EFL Saudi students majoring in English Language from the Department of English Language at the College of Arts, King Faisal University, Saudi Arabia were randomly selected through a computer randomizer. They were enrolled in a four-year bachelor's degree in English Language. Each year, students study a pool of core and optional courses in English language, Linguistics, Translation and Literature, over two semesters. The chosen participants were uniformly aged between 18 and 21 years, sharing similar English language backgrounds. Their academic levels were ascertained based on comprehensive evaluations, including tests, quizzes, and class participation, conducted throughout the two-year period. It is important to highlight that the role of language proficiency in relation to Test-Taking Strategies (TTS) remains a subject of ongoing debate. The inclusion of female students was impractical for several reasons, such as cultural constraints, the nature of the TA protocols (which necessitate the full presence of the researcher), and challenges related to access. Furthermore, previous research in the Saudi context has shown that gender was not a significant factor in relation to TTS (Al Fraidan, 2011).

Data collection and procedures

Data collection lasted for two years, beginning in their second year and ended in the last semester before graduation. In the first year, the data collection took place through think-aloud protocols and semi-structured interviews, while taking different teacher-made achievement tests in the "Study Skills" course in the first semester and in the "Building Vocabulary" course in the second semester. In the second year, courses included Reading Comprehension" and "Critical Reading. The researcher opted to focus on courses he personally taught to facilitate and streamline the data collection process. All tests gauged their vocabulary knowledge from the list of words in each course using multiple-choice gap-filling test formats. The selection of multiple-choice format is attributed to its widespread usage in Saudi Arabia (Addamegh, 2011; Al Fraidan, 2010). Furthermore, the distinctive characteristics of this format often prompt various TTS.

The longitudinal method was employed for several reasons. Firstly, it was chosen for a more profound and careful analysis of TTS. Secondly, the researcher aimed to identify valid instances of TTS used by the participants to ensure a reliable analysis, eliminating the capture of a single occurrence of TTS, and providing a more realistic picture of the strategic competence of the participants. Thirdly, the method was utilized to explore the evolution and development of TTS over time. However, it's important to note that this study did not focus on investigating the evolution and development of TTS, leaving it as a potential future line of inquiry.

The students were trained in performing think-aloud through various non-linguistic tasks, such as solving mathematical problems. Subsequently, individual verbalization sessions were conducted with each participant, followed by semi-structured interviews. The interviews primarily centered on obtaining further explanations regarding the use of TTS and the underlying reasons behind the selection of specific strategies or answers.

Transcriptions, translations, and transliterations of verbal protocols took place one month after each test. A total of eight tests were conducted. The analysis and coding of the TTS took two consecutive months and included validity and reliability checks by four faculty members from different universities, who were either independent associate or full professor. Each faculty member was provided with the same protocol to code TTS, ensuring consistency and robustness of the analysis. Owing to discrepancies in coding for certain TTS, a decision was made to exclude a particular TTS from the analysis if there was disagreement among three coders. The researcher opted for a longitudinal approach to authentically capture the utilization of TTS. This choice aimed to go beyond students merely expressing their strategy preferences and to ensure an accurate assessment of the actual frequency of TTS usage over time.

The focus of the analysis was on identifying multiple instances to demonstrate that TTS can serve multiple functions, manifesting in clusters and being operationalized both wisely and unwisely. Each TTS was assigned two codes, indicating its function and the wiseness of its operationalization. Subsequently, TTS clusters were determined based on the number of instances used to guess or achieve an answer.

IV. RESULTS

The main focus of this research was to provide more insights into the three neglected characteristics of LLS in general and TTS in particular. These are strategy clustering, multifunctionality of strategies, and operationalization.

Strategies clustering

It was already found that a single strategy cannot be used alone to complete a language task (Graham, 1997; Neubach & Cohen, 1988; Laviosa, 2000; Macaro, 2001, 2003, 2006; Addamegh, 2003; Phakiti, 2006; Al Fraidan, 2019; Cohen et al., 2023a). This study aimed to further investigate the nature of the resultant clustering of TTS. For instance, when a student begins tackling a test item by re-reading the question, repeating key words, inferring the meaning of the gapped word, testing a potential candidate word, inserting the candidate word in the gap, reviewing the entire sentence with the candidate word to assess its fit, and reaffirming the entire sentence or candidate word for confirmation. The researcher identified 92 different strategy types from eight test occasions. They were classified into three main categories: TTS before attempting the task/question, TTS while tackling the task/question, and TTS after completing the task/question. Due to space constraints in this article, the researcher highlights the top four most frequently used TTS in each of the three main categories. This selection, totaling 12 TTS, serves as illustrative examples to elucidate the arguments

presented here (Table 1). The frequency of TTS was calculated according to the token/type ratio and its appearance on more than two test occasions.

TABLE 1
THE THREE MAIN CATEGORIES OF TTS WITH EXAMPLES OF THE TOP FOUR MOST FREQUENTLY USED ONES

A. TTS before attempting the task/question	
1.	Reciting religious phrases to minimize anxiety.
2.	Getting acquainted with the nature of the task (test formats, number of questions, allocated time and scores)
3.	Deciding the order of attempting questions (by easiness or difficulty)
4.	Asking for more clarifications about test instructions/task conditions
B. TTS while tackling the task/question	
1.	Inferencing the meaning of the gapped word
2.	Inferencing the part of speech of the gapped word
3.	Eliminating possible distractors
4.	Translating the question from L2 into L1
C. TTS after completing the task/question	
1.	Revising the answer by checking other possible distractors
2.	Confirming all answers or suspected answers after finishing the whole test
3.	Changing a confirmed answer for different reasons (hesitation, bias on the test)
4.	Paraphrasing/ translating the full sentence with the selected answer

Table 2 shows descriptive analysis of the three main categories. It signals that category B is the main pool of the strategies, which was expected as this is where students try to struggle with finding an answer, facilitating a problem or compensating a deficit in their lexical knowledge.

TABLE 2
DESCRIPTIVE ANALYSIS OF THE THREE MAIN CATEGORIES

TTS category	Me an	Std. Deviation
TTS A	2.1 875	2.22767
TTS B	6.2 500	3.49285
TTS C	.25 00	.57735
N= 20		

The data passed the Kolmogorov-Smirnov test of normality; and the suitability of using the ANOVA test showed a significantly strong interaction between the three categories ($F=25,000$, $p=.000$). It illustrates how each category serves distinct functions in accomplishing a test task. Furthermore, the correlation analysis among the three categories was highly positive (Table 3). This can be considered as the main cluster required to complete any test task. The distribution of strategies under these three categories is not exclusive, as some strategies fall into more than one category (translating the questions from L2 to L1 can be in B or C). This is confirming the multifunctionality of the strategies, which will be discussed later.

TABLE 3
CORRELATIONS AMONG THE THREE CATEGORIES

Category Pair	Correlation	
TTS A/ TTS B	Pearson correlation	.808**
	Sig. (2 tailed)	<.001
TTS B/ TTS C	Pearson correlation	.895**
	Sig. (2 tailed)	<.001
TTS A/ TTS B	Pearson correlation	.814**
	Sig. (2 tailed)	<.001

The researcher examined the detail of TTS to identify frequent clustering among them. This deep analysis revealed different frequent clustering patterns among main categories. For example,

Cluster 1: A1, B1, C2

Cluster 2: A2, B3, C1

Cluster 3: A3, B4, C3

These examples only show clusters in three strategies. The remaining clusters contained 21 strategies. The analysis revealed several such clustering strategies, which could be further investigated to check different aspects such as interactivity between strategies within each category and cluster, and the efficiency or impact of each cluster on the answers. However, the researcher thought the patterns of clustering itself would shed more light on how some of the strategies clustered to complete a task. For example, one student used twelve strategies to answer a test item.

The general manager decided to give a to his secretary.

- *professional*

- ***promotion***
- ***productivity***
- ***prosperity***

The student attempted the question using 12 strategies in this order: rereading the stem more than once, rereading the distractors several times, attempting each distractor in the gap, eliminating possible distractors, questioning the possibility of more than one distractor, inferencing the meaning of the gap, eliminating distractors based on semantic knowledge, inferencing the part of speech of the gap, checking how the selected answer feels in the sentence, translating the full sentence from L2 into L1, changing the previously selected answer “promotion” to the new “prosperity,” confirming the answer, and then moving to another question. This example illustrates the logical clustering of strategies utilizing different types of knowledge (semantic and syntactic). It was one of the most frequently used clusters among participants. Other strategy clustering showed the illogical progression of strategies, such as identifying the meaning of a gapped word and then choosing an answer based on its weirdness. An example of a think-aloud protocol is as follows: “mmm.. This sentence needs a word that means a bonus.. mmmm this word “prosperity” looks so weird let me put it”.

This can be further explored to gain insight into illogical clustering and the underlying reasons behind it. This observation aligns with the earlier discussion of the three stages of the PPP model, emphasizing how diverse interactions of strategies within these stages contribute to distinct strategy clusters. The study highlights dual interactions, which, at times, result in the selection of the correct answer and, conversely, may lead to changing the initially correct choice to an incorrect one.

Multifunctionality of strategies

While it has long been established that TTS can be categorized based on their functions (cognitive, metacognitive, social, and affective), a central assertion of this study is that these strategies are inherently multi-functional and multipurpose. Consequently, classifying them into a single functional category is deemed illogical. Cohen et al. (2023a) propose the structure shown in Figure 2 to prove that strategy functions fluctuate. One strategy for determining the meaning of a word in a context can have four functions.

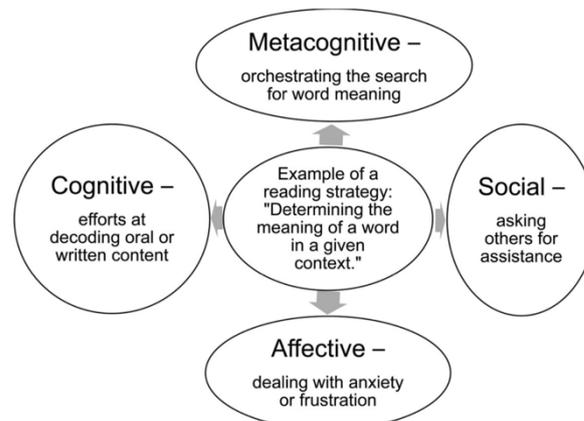


Figure 2. Fluctuation in the Functions of Any Given Strategy in the Performance of a Given Task (Cohen et al., 2023a)

This argument was proven in this study. However, this study adds a new characteristic which is multifunctionality. Each of these four functions can be used to serve different purposes. This study categorizes three main purposes (facilitating, compensatory, and problem solving). A strategy is facilitative when a student wants to overcome the difficulties faced while attempting a test item, such as rereading the test item or stem. The compensatory aspect of strategies occurs when a student possesses partial knowledge of the tested item, such as using different retrieval strategies for knowing the meaning of a gapped word, but not recognizing its form. A strategy is considered problem-solving when the student finds that the test item is very difficult for different reasons, such as skipping for not fully knowing the test item or facing problems with understating the test item/instructions. An illustration of this phenomenon occurs when a student repeats a key word in a test item. This repetition can serve as a facilitative approach to reaching an answer without encountering difficulties or struggling with the test item. Alternatively, it may be viewed as a method to compensate for a partially forgotten aspect of the gapped word, addressing a deficit in lexical knowledge. Moreover, this repetition can be seen as a strategy to solve the problem of the overall comprehension of the test item.

In previous studies, such as Addamegh (2006), Alyami (2011), Khan and Rasheed (2019), Killi and Leu (2019), Phakiti (2003), classifications of TTS were organized into reasonable and logical categories. However, these studies often overlooked the possibility that a strategy could serve multiple functions and roles. The current study challenges this notion, revealing that strategies indeed assume different roles and various functions within distinct clusters. Future

analyses exploring the correlation and interaction between these clusters and their intended purposes will enhance our comprehension of the dynamic nature of TTS.

TTS operationalization

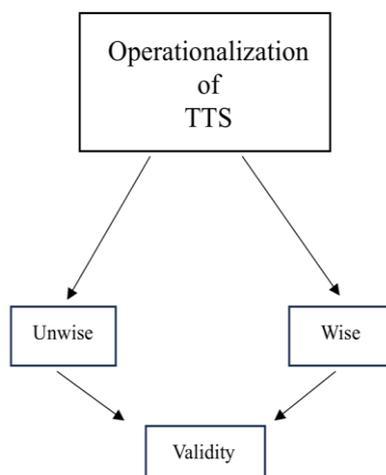


Figure 3. Operationalization of TTS

TTS were used in test situations with different degrees of wiseness by students. For example, in one test question, they may start by using test-wiseness strategies such as blind guessing but then finish with construct-relevant strategies, showing their true knowledge of the tested construct. In most cases, there is a mixture of wise and unwise operationalization. That is, a student can wisely operationalize a strategy or cluster of strategies to find the correct answer, or vice versa, sometimes changing the correct answer to an incorrect one. Figure 3 shows the important issue of how the operationalization of TTS can affect test validity. This study found that there are multiple instances of wise TTS that secures correct answers without processing the construct or possessing the required knowledge, which negatively affects the validity of the test scores. This occurred through wild guessing and the utilization of bias in the test. The unwise operationalization of the TTS can also affect test validity, when students change their answers from correct to wrong. It was found that 21% of test items had their answers changed from correct to wrong answers, among all the tests. This change was considered only when the student wrote an answer on the test paper, and then visibly crossed it off to write another answer. This behavior was found more frequently among high-level students (six students and three low-level students).

Students tended to experience fluctuating correct answers with incorrect answers, despite knowing the correct answers. This confusion occurs when students cannot distinguish between the alternatives.

The think-aloud protocols analysis revealed that verbalizing terms incorrectly caused students to alter their correct answers to incorrect ones. To illustrate, the word "sack," was verbalized "sick" leading students to choose the wrong answer or alter a correct one into wrong. Consequently, students' scores do not reflect their knowledge of the intended word. This would be considered deceptive if it is applied in high-level tasks that have life-altering consequences, such as placement and admission entry tests (Yang et al., 2019).

One reason for test-unwise strategies are test flaws such as repeating the same word, which might provide a hint to students or applying a surface match between the tested word and the choices. Another reason is that students roughly guess unknown answers without processing any relevant test constructs or meanings of the questions, for various hidden reasons, including their laziness. According to Al Fraidan and Al-Khalaf (2012), students sometimes do not want to spend time thinking of test items, as they lack the required knowledge to provide correct answers due to test anxiety or not having studied well for the tests.

A third reason is that participants did not read the instructions carefully. In this case, while attempting the tests, students tend to rush to use their previous experience or memory of standard tests instead of focusing on the actual test (Sagar, 2019). For example, when students answer a cloze test by inserting words from the text itself, they think that the right way to solve a cloze test is to find words from the text itself. In addition, the lack of clarity in test instructions can also result in poor behavior. Consuming more time with challenging items while simply skipping them and answering the easiest ones is another source of unwise strategies (Widyantoro, 2019).

The above findings can contribute valuable insights to test design. Instructors must recognize the significance of providing clear test instructions and ensure the accuracy of test items (Al Fraidan & Al-Khalaf, 2012). It is imperative for teachers to prioritize the meticulous crafting of tests, avoiding any potential confusion resulting from flaws in the test, as these can adversely impact the validity of the scores.

V. DISCUSSION

Very often, an individual strategy, or a pair or cluster of strategies is the primary cause of success or failure for a given assessment item or task (Cohen et al., 2023b). Learners may sometimes obtain unfortunate results owing to a lack of awareness of how to effectively utilize TTS. The implementation of TTS is affected by different factors, such as students' background knowledge, proficiency, test design, and strategic competence. Some of these factors may lead to negative results such as invalidity of the results and test anxiety.

TTS is examined in a manner that enables students to avoid the development of interventions by boosting positive skills aimed at improving their academic performance (Sagar, 2019). In most cases, the test is expected to assess students' knowledge of different disciplines. However, with other factors coming into play, this could negatively or positively affect students' performance. The latter affects learner performance by rendering test scores ineffective (Kivity & Huppert, 2016). In addition, test-taking strategies become capable of increasing test score validity from a general perspective, thus enabling students to accurately reflect on what has been learned in the classroom and what they can comprehend (Flippo et al., 2018).

This study offers a comprehensive examination of TTS from various perspectives, leading to the introduction of a novel definition. The development of this new definition stems from critiques of previous definitions and classifications, which were deemed insufficient in capturing crucial functions and roles. While it is acknowledged that strategies cluster together to reach an answer, the focal point lies in understanding how these strategies are operationalized—whether wisely or unwisely. Additionally, the study emphasizes the importance of scrutinizing the functioning of TTS during a test and identifying the specific roles they assume, whether facilitative, compensatory, or problem-solving. Furthermore, it is essential to highlight that test-taking techniques do not have a limit to their implementation, as they fit within any content and institutional level. By engaging both students and instructors in understanding the complexities of how TTS is orchestrated and operationalized, it is easier for both groups to reach valid evaluations (Yang et al., 2019).

As a result, different research and academic methodologies should be implemented to enhance students' confidence, avoiding scenarios where they delete their correct answers before submitting the test. It is essential to note that poor performance in EFL tests is mostly attributed to poor management of time, carelessness, and confusion caused by unknown strategies and item formats. Therefore, as test scores are affected by test-wiseness, it is important for individuals involved in their administration, development, and interpretation to critically consider the construct of test-wiseness and how it affects students' scores (Kornejady et al., 2019).

VI. CONCLUSION

One important contribution of this study is how it enhances the definition of TTS by adding three important characteristics: clustering, multifunctionality, and the degree of wise operationalization. The study further introduces the concept that TTS can assume three distinct roles in aiding the selection of an answer—namely, facilitative, compensatory, or problem-solving roles. Notably, these three characteristics have recently been identified as actual test behaviors. Despite their significance, researchers have not dedicated adequate attention to exploring these roles comprehensively.

This direction should be taken for further analysis, with different contexts, types of learners, and test formats. This longitudinal study was limited to one context: learner type, language skill (reading), and language component (vocabulary). Future studies can expand the scope of these factors, to produce more multifaceted analyses.

There is a need for an additional examination of how TTS is operationalized, since students perform different TTS clusters and with different degrees of wiseness of operationalization. In addition, another direction that was found in this study but could not be completely confirmed is that the relationship between various TTS clusters, as one cluster can lead to another cluster. In addition, the interactivity between the three characteristics and how all of these can affect test validity can also be explored further.

The think-aloud approach is a time-consuming tool with lots of efforts to be made in order to reach findings including transcriptions, translations, transliterations, and coding. It is highly recommended to start to improve this tool with Artificial Intelligence technology to help reduce time and efforts.

Lastly, Educators need to investigate test-unwise strategies in particular and offer more insights into why and how they occur. They are also required to provide practical illustrations to students to help them understand how they score lesser if they operationalize these TTS with different degrees of wiseness. The variance between test-unwise and test-wise students could be due to differences in cognitive monitoring.

Test-taking strategies coaching plays a pivotal role in empowering individuals to navigate assessments effectively. The ability to employ TTS during testing situations can significantly impact validity and reliability. Such coaching not only familiarizes individuals with various types of test formats but also imparts valuable insights into time management, question interpretation, and stress reduction. It equips learners with a toolkit of techniques, ranging from effective problem-solving methods to efficient resource allocation, fostering confidence and resilience in the face of challenging exams. Students need to understand that TTS work at cross-purposes helping them to be aware of how to operationalize the test wisely. Moreover, test-taking strategies coaching instills a deeper understanding of one's cognitive processes, promoting metacognition and enhancing overall learning outcomes. As assessments increasingly become gatekeepers for educational and professional opportunities, the importance of imparting effective test-taking strategies becomes instrumental in shaping individuals into adept and successful test-takers.

ACKNOWLEDGEMENTS

I would like to acknowledge all the people who facilitated this project including administrators, faculty members, family members and the research participants for their cooperation in this project. Special thanks to my lovely wife, AlAnoud Alwasmī for facilitating producing this research.

FUNDING

This work was funded and supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia [Grant No. 5,879].

REFERENCES

- [1] Addamegh, K. (2003). *EFL multiple choice vocabulary test-taking strategies and construct validity* [Unpublished doctoral thesis]. University of Essex.
- [2] Al Fraidan, A. (2011). *Test-taking strategies of EFL learners on two vocabulary tests*. Lap Lambert Publications. ISBN 978-3-8454-7030-6.
- [3] Al Fraidan, A., & Al-Khalaf, K. (2012). Test-taking strategies of Arab EFL learners on multiple choice tests. *International Education Studies*, 5(4), 80-85. doi:10.5539/ies.v5n4p80
- [4] Al Fraidan, A. (2014). Test-unwiseness: what are they? *Journal of Applied Sciences*, 14, 8. doi:10.3923/jas.2014.828.832
- [5] Al Fraidan, A. (2019). New test-taking patterns and their effect on language test validity. *Language Testing in Asia*, 9(11), 1-28. doi:10.1186/s40468-019-0088-5
- [6] Al Fraidan, A., & AlSalman, A. (2023). The utilization of test-taking strategies by high schoolers in Saudi Arabia. *World English language Journal*, 13(2), 414-423. doi:10.5430/wjel.v13n2p414
- [7] Allan, A., & MacLennan, C. (1997). Validating a test of test-taking strategies for use in Hong Kong secondary schools. *Perspectives*, 9, 1-5.
- [8] Alsafi, W. A. (2019). Using Verbal Reports to Explore Reading Test-Taking Strategies of Saudi EFL University Students. *Arab World English Journal, Special Issue 1: Application of Global ELT Practices in Saudi Arabia*, 42-56. doi: https://dx.doi.org/10.24093/awej/elt1.4
- [9] Alyami, S. (2011). *Vocabulary learning strategies of Saudi EFL majors of different gender, year and proficiency: use and reasons for use* [Unpublished doctoral thesis]. University of Essex.
- [10] Baker, L., & Brown, A.L. (1984). Metacognitive skills and reading. In P.D. Pearson, R. Barr, M.L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 353-394). Longman.
- [11] Chamot, A. (2005). Language learning strategy instruction: Current issues and research. *Annual Review of Applied Linguistics*, 25, 112-130.
- [12] Chamot, A., Barnhardt, S., El-Dinary, P., & Robbins, J. (1996). Methods for teaching learner strategies in the foreign language classroom. In R.L. Oxford (Ed.), *Language learning strategies around the world: Cross-cultural perspectives* (pp. 175-187). University of Hawaii, Honolulu: Second Language Teaching & Curriculum Center.
- [13] Chamot, A., & Kupper, L. (1989). Learning strategies in foreign language instruction. *Foreign Language Annals*, 22, 13-24.
- [14] Cohen, A. (2013). Using test-wisness strategy research in task development. In A. J. Kunnan (Ed.), *The companion to language assessment – Vol. 2: Approaches and development, Part 7: Assessment development* (pp. 1-12). Wiley and Sons.
- [15] Cohen, A. (2011). *Strategies in learning and using a second language* (2nd ed.). Harlow Pearson/Longman.
- [16] Cohen, A., & Upton, T. (2007). "I want to go back to the text": Response strategies on the reading subtest of the new TOEFL. *Language Testing*, 24(2), 209-250.
- [17] Cohen, A., & Upton, T. A. (2006). *Strategies in responding to the new TOEFL reading tasks* (Monograph No.33). Princeton: Educational testing service.
- [18] Cohen, A. D., Gu, P. Y., Nyikos, M., Plonsky, L., Harris, V., Gunning, P., Wang, I. K., Pawlak, M., Gavriilidou, Z., Mitits, L., Sykes, J. M., & Ga, A. (2023a). Tangible insights on the strategizing of language learners and users. *Language Teaching*, 56, 313–332. doi:10.1017/S0261444823000046
- [19] Cohen, A. D., Rahmati, T., & Sadeghi, K. (2023b). Test-Taking Strategies in Technology-Assisted Language Assessment. In K. Sadeghi & D. Davis (Eds.), *Fundamental Considerations in Technology Mediated Language Assessment* (pp. 235–254). Routledge.
- [20] Cohen, A. D. (1994). *Assessing language ability in the classroom* (2nd ed.). Heinle & Heinle.
- [21] Cohen, A. D. (1998). *Strategies in learning and using a second language*. Longman.
- [22] Cohen, A. D., & Pinilla-Herrera, A. (2009). Communicating grammatically: Constructing a learner strategies website for Spanish. In 2009 LTTC International Conference on English Language Teaching and Testing. *A new look at language teaching and testing: English as subject and vehicle* (pp. 62-74). Taipei, Taiwan: Language Teaching and Testing Centre.
- [23] Flavell, J. H. (1981). Cognitive monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills* (pp. 35-60). Academic Press.
- [24] Graham, S. (1997). Effective language learning: Positive strategies for advanced level language learning. *Modern Language in Practice*, 6, 216-225.
- [25] González, D. (2007). Text-to-speech applications used in EFL contexts to enhance pronunciation. *TESL-EJ*, 11(2), 1-11.
- [26] Han, F. (2023). Test-taking strategy use by Chinese young EFL learners in an authentic international standardized English test. *Porta Linguarum an International Journal of Foreign Language Teaching and Learning*, 39, 265–280. https://doi.org/10.30827/portalin.vi39.24390
- [27] Hirano, K. (2009). Research on test-taking strategies in L2 reading. *Bull. Joetsu Uni. Educ.*, 28, 157-165.

- [28] Hosenfeld, C. (1984). Case studies of ninth graders. In J.C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 231-249). London: Longman.
- [29] Khan, M., & Rasheed, S. (2019). Moderating Role of Learning Strategies Between Meta-Cognitive Awareness and Study Habits among University Students. *Pakistan Journal of Psychological Research*, 34(1), 215-231. <https://doi.org/10.33824/PJPR.2019.34.1.12>.
- [30] Kiili, C., & Leu, D. J. (2019). Exploring the collaborative synthesis of information during online reading. *Computers in Human Behavior*, 95, 146-157. doi:10.1016/j.chb.2019.01.033.
- [31] Kornejady, A., Pourghasemi, H. R., & Afzali, S. F. (2019). Presentation of RFFR new ensemble model for landslide susceptibility assessment in Iran. In *Landslides: Theory, Practice, and Modelling* (pp. 123-143). Springer, Cham.
- [32] Laviosa, F. (2000). The listening comprehension processes and strategies of learners of Italian: A case study. *Rassegna italiana di linguistica applicata*, 32(2), 129-159.
- [33] Macaro, E. (2001). *Learning strategies in foreign and second language classrooms*. London: Continuum.
- [34] Macaro, E. (2003). *Teaching and learning a second language: A guide to current research and its applications*. London: Continuum.
- [35] Macaro, E. (2006). Strategies for language learning and for language use: revising the theoretical framework. *Modern Language Journal*, 90(3), 320-337.
- [36] Neubach, A., & Cohen, A. D. (1988). Processing strategies and problems encountered in the use of dictionaries. *Dictionaries: Journal of the Dictionary Society of North America*, 10, 1-19.
- [37] Nikolov, M. (2006). Test-taking strategies of 12-13-year-old Hungarian learners of EFL: Why whales have migraine. *Language Learning*, 57(1), 1-51.
- [38] Nourdad, N., & Ajideh, P. (2019). On The Relationship between Test-Taking Strategies and EFL Reading Performance. *Journal of English language*, 11(23), 189-219.
- [39] Oxford, R. (2010). *Teaching & Researching Language Learning Strategies*. Longman.
- [40] Oxford, R. L. (1990). *Language Learning Strategies: What Every Teacher Should Know*. Heinle & Heinle.
- [41] Rogers, W. T., & Harley, D. (1999). An empirical comparison of three-and four-choice items and tests: susceptibility to testwiseness and internal consistency reliability. *Educational and Psychological Measurement*, 59(2), 234-247.
- [42] Rupp, A. A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: a cognitive processing perspective. *Language Testing*, 23(4), 441-474.
- [43] Sagar, N. (2019). *Exploring a combined instructional and self-regulatory intervention for math anxiety*. [Doctoral dissertation, University of British Columbia].
- [44] Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Descriptive, acquisition and pedagogy* (pp. 195-215). Cambridge University Press.
- [45] Widyantoro, A. (2019). Teaching Conditional Sentence Using Presentation, Practice, and Production (PPP) Technique in Vocational School. *CLIENT (Culture, Literature, Linguistics, English Teaching)*, 1(02), 147-164.
- [46] Yang, B., Wei, J., Ju, P., & Chen, J. (2019). Effects of regulating intestinal microbiota on anxiety symptoms: a systematic review. *General psychiatry*, 32(2)1-9.



Abdullah A. Al Fraidan is an associate professor of applied linguistics at the Department of English Language, College of Arts, King Faisal University, Saudi Arabia. He was born in ALHOFUF in 1979. He is a former Dean of eLearning and Distance Education, responsible of more than 150,000 students across the world studying BA degrees in different subjects by online distance. He was also a former head of the English Language Department and a former Director of the English Language centre. He published more than 20 publications in language testing, language technology, sociolinguistics, program evaluation, discourse analysis, lexicology, social media use in assessment, Mobile Assisted Language Learning and lexical attrition. He is the recipient of the 2016 British Council Alumni Award as the best Saudi Alumni for social impact category.