**LLT Journal**, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 29, No. 1, April 2026, pp. 381-400

*LLT Journal: A Journal on Language and Language Teaching*
http://ejournal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

# GUTTMAN CHART FOR LANGUAGE TEST SCORES ANALYSIS

**Mohamed El zerk**
Mohammed V University, Morocco
mohamed.elzerk@um5r.ac.ma
correspondence: mohamed.elzerk@um5r.ac.ma
https://doi.org/10.24071/llt.v29i1.379
received 01 August 2024; accepted 08 March 2026

## Abstract
Objective discrete-point language tests are widely used across various EFL contexts, including Morocco, because of their practicality, neutrality and precision. Therefore, it is crucial to explore effective methods for systematically analysing test scores to guide formative instructional decisions and improve test item quality, given the pervasive use of their results for high-stakes decisions, such as accountability and placement. To this end, this study investigates the use of the Guttman chart to analyse an objective test, identify a group of high school Common Core Science students' Zone of Proximal Development (ZPD), and develop a suitable remedial. Additionally, it assesses the performance of test items at both the general and individual levels. Employing a repeated-measures pre-test and post-test design, this study examines the impact of the remedial work, based on a Guttman chart analysis of a written pre-test, on students' performance in an online post-test. The findings demonstrate that the Guttman chart is effective for analysing objective test scores, facilitating formative decision-making, and evaluating test item functionality. This is supported by the statistically significant improvement in students' scores from the pre-test to the post-test. Furthermore, the study offers valuable implications for applying the Guttman chart in test score analysis.

**Keywords:** Guttman chart analysis, remedial work, scores, test items

## Introduction
The field of language testing has undergone a significant evolution, with increasing recognition of the uniqueness of the classroom context and the teacher's role in interpreting assessment (Turner, 2012). This is because of the complex nature of the classroom as a social environment (Leighton, 2019) and the difficulty of accounting for individual student differences in designing and using assessments (McMillan, 2019). Therefore, the accuracy of interpreting assessment results locally is affected (Ferrara et al., 2019). Given the practicality and feasibility of their design, administration, and scoring (Giraldo, 2020), classroom objective tests constitute the predominant assessment tool for measuring students' progress or achievement relative to a specific sampled domain. This is especially significant in the Moroccan EFL context (Benzehaf, 2017; El zerk, 2025a; El zerk & Ichebah, 2025; Mamad & Vigh, 2021), where five formal written quizzes and a final global test are administered per semester (NCEEO, 2007).

Despite the increasing debate from a critical perspective regarding the negative effects of objective testing on teaching and learning, such as curriculum narrowing, an emphasis on test-taking strategies, and other unintended psychological and motivational impacts on students (Tsagari & Cheng, 2017; Xie & Jia, 2025), Fox (2016) and Giraldo (2020) provide counter-evidence in favor of discrete-point tests. These are considered reliable tools for efficient decision-making that meet statistical requirements, faithfully represent language constructs and support meaningful instructional decision-making. Additionally, language tests offer information about a sample of items related to, for instance, vocabulary and grammar (Bailey, 2016), which demonstrate student mid-unit progress to adjust instruction accordingly (Bonner, 2013). However, Ismail and Yoestara (2017) found that EFL teachers in Aceh, Indonesia, rarely conduct the index-determining step or tryouts when constructing a test, undermining the items' quality. This reinforces the limited understanding of how to improve EFL teacher-made evaluations (Ghaicha, 2016), especially within Morocco's high-stakes testing context.

Valid and reliable test results from continuous assessment can inform multi-level groupings and help determine the type of instructional materials that need to be covered (Nouri, 2021). They can also inform the design of instruction and guide the reform of the curriculum (Fan et al., 2021). As Brookhart (2019) notes, "any language test has some potential for providing diagnostic information" (p. 60). Following this perspective, the teacher assumes a pivotal role in interpreting test scores to identify the gap between current and expected performance (Schneider et al., 2013). This ongoing formative use of assessment data enables teachers to determine students' current developmental stages through the provision of ongoing constructive feedback and planning challenging tasks accompanied by scaffolding (Anwar et al., 2024). Markedly, this practice is especially necessary given the widespread use of tests and quizzes in various high-stakes educational contexts, such as Morocco, to establish a balanced assessment design.

In the Moroccan EFL context, discrete-point quizzes and tests featuring objective items such as multiple choice and gap-fill are the predominant formal techniques for assessing EFL components (El zerk, 2025a; El zerk & Ichebah, 2025; Mamad & Vigh, 2021). Besides summarizing students' achievement, teachers can also use test scores, which, when analysed adequately, could inform remedial teaching programs meant for differentiation and scaffolding (Nouri, 2021). However, there is a lack of explicit guidelines on the systematic implementation of formative assessment for effective intervention (e.g., differentiation and lesson adjustment) (van der Steen et al., 2022). Specifically, how teachers incorporate test information to guide their teaching and enhance students' learning remains a significant research gap (Jang, 2017). On this point, the formative use of assessment to determine students' current developmental stages and provide constructive feedback facilitates curriculum design and lesson planning within the framework of the ZPD (Anwar et al., 2024).

Intended to address the gap in developing high-quality test items and to ensure the use of test scores for formative decision making, this study is consistent with the functional perspective that emphasizes the role of assessment in improving instructional quality and learning for all students (Kane & Wools, 2019; Tomlinson

& Moon, 2013). This study is also grounded within the framework of ZPD, which informs planning differentiated instruction and scaffolding (Anwar et al., 2024; Wang & Feng, 2023). Thus, it examines the impact of applying Guttman chart analysis to objective EFL test scores in facilitating the planning of a remedial work designed to address students' identified weaknesses. Additionally, it evaluates the functionality of test items at both general and specific levels. According to the author's knowledge, recent studies using the Guttman chart for language test scores analysis are non-existent.

Therefore, this study promises significant implications for EFL teachers and test designers, suggesting the adoption of a systematic and evidence-based approach to differentiation. Importantly, this approach ensures that instruction is tailored to students' readiness levels, drawing on objective test scores; thereby, it maximizes learning potential and improves test-item quality, especially in high-stakes testing contexts. For this reason, this study addresses the following research questions:

1. Does Guttman chart analysis of the objective EFL test adequately inform the planning of remedial work?
2. Do all items measure the intended EFL components accurately?
3. Is there a significant difference between students' pre-test and post-test scores due to the remedial work?

## Literature Review

This section lays a comprehensive foundation for the relevant constructs of this study. It provides an overview of language testing, emphasising the EFL context, the Zone of Proximal Development (ZPD), and the Guttman chart analysis, including the design process. A central construct is the Guttman chart, defined in this study as an organized spreadsheet that presents data of an objective test in an easily interpretable visual format to identify students' ZPD.

### *Reframing EFL testing for formative purposes*

In EFL educational settings, teachers spend considerable time assessing their students' knowledge of language relative to an instructional program's learning objectives for a variety of purposes, including checking their progress for future instructional decision-making or measuring their achievement for grading purposes. One pervasive assessment tool is a test, which is defined as "any systematically administered set of procedures that results in some level of measurement of an attribute" (Bonner, 2013, p. 88). Assessment tools, including tests, provide valuable data that serve different purposes, whether formative (e.g., plan instruction), summative (e.g., certification, selection), or a combination of both (El zerk, 2025b). Despite being relatively low-stakes, progress tests and quizzes could provide valuable diagnostic information and contribute to informing the teaching process. In this regard, a central question remains: "how teachers can best design and implement formative assessment to have it really contribute to better or better-founded formative decision-making" (van der Steen et al., 2022, p. 20).

A prominent function of language tests that ensures positive washback on teaching and learning is to interpret their scores for inferring diagnostic information about the measured attributes, accompanied by an action plan to enact positive change. This is necessary to address what Nouri (2021) named *Indirect underachievement*, a case where students pass the test with achievement gaps or deficits that may become fossilized later if not remediated. To fulfil this function, Ismail and Yoestara (2017) stress that teachers are required to devise high-quality tests that provide information on students' mastery levels in relation to the materials covered to inform decision-making. High-quality tests and items also ensure confidence in high-stakes decisions that rely primarily on their scores.

Language tests remain a primary formal assessment technique to evaluate students' knowledge and skills (Widyastuti & Harto Pramono, 2021). The attributes measured, such as grammatical knowledge are not directly observable features. Hence, a construct is a precise definition of an attribute for a particular measurement situation, incorporating different activities (Bachman, 2004). In terms of items, Randel and Clark (2013) argue that tests feature Selected Response (SR) (i.e., multiple choice) and Constructed-Response (CR) (i.e., short answer and essay). Objective language tests are characterized by the predominant use of SR items. The quality of teacher-made SR test items rests on the clarity of the learning objectives, the teacher's subject matter expertise and their skills in item writing (Rodriguez & Haladyna, 2013). Furthermore, Bachman (2004) stresses the need to demonstrate the reliability of the language test scores. Nevertheless, Ismail and Yoestara (2017) found that EFL teachers barely conduct item analysis when constructing tests. This emphasizes the necessity to identify alternative practical analysis procedures.

It should be noted that one of the main limitations of these objective tests is that they do not explain the reasons behind students' performance (Russell, 2019). However, when adequately analyzed, they may serve as primary sources of numerical data within a comprehensive assessment process that incorporates both quantitative and qualitative data-gathering procedures. Regardless of the techniques employed, Settiawana and Hilmawan (2016) conceptualize assessment as a cycle consisting of planning, gathering evidence, interpreting evidence, and using the evidence. Notably, interpreting and using assessment evidence (e.g., test scores) for instructional decision-making constitutes a crucial stage of the assessment process. On this basis, Kane and Wools (2019) emphasize the accuracy of assessment by framing teachers' data interpretation within students' current levels of achievement and their learning profiles. Adequate interpretation of test scores will identify areas where students need targeted instruction or remedial work, which can be adequately understood through the concept of ZPD.

***The Zone of Proximal Development (ZPD) as a framework for remedial work***

A salient purpose behind interpreting students' test scores is to determine their ZPD. Conceptually, Vygotsky (1978) argues that ZPD defines the "functions that have not yet matured but are in the process of maturation" (p. 86). In other words, the ones that are still in an embryonic state (Vygotsky, 1978). Furthermore, his notion of ZPD is based on his hypothesis that a child's developmental process does not coincide with, but rather lags behind, the learning process. This implies the necessity of providing children with sufficient assistance to progress from their current level of competence to realize their potential for learning. By extending this model to language assessment, one can account for the learning achievements that have been developed as well as the ones that are still in the process of development based on three levels: the actual developmental level, the zone of proximal development and the zone of future development (see Figure 1). Hence, assessment and teaching are integrated, with assessment results (e.g., test scores) and other information playing a significant role in informing formative instructional decision-making.

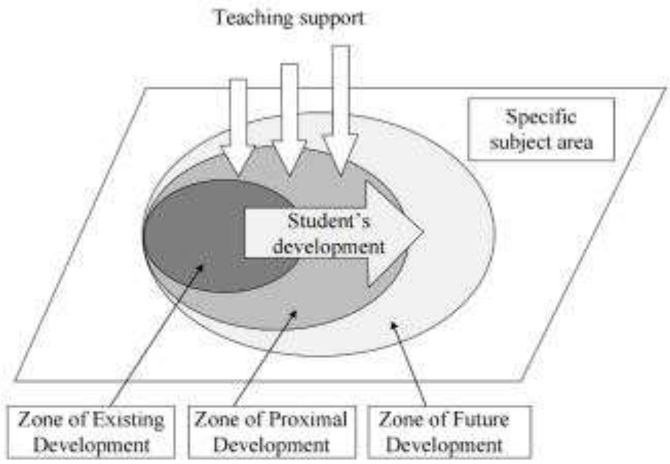


Figure 1. ZPD framework for developing remedial work (Wang & Feng, 2023, p. 45)

The implementation of the concept of ZPD for developing remedial work has expanded to include a wide range of scaffolding activities that students cannot complete independently (Zaretsky, 2024). Drawing on this, teaching is viewed as part of assessment regardless of its purpose or context by revealing emerging and developing abilities, which is known as dynamic assessment (Poehner et al., 2017). It is advantageous in that it focuses on the learning process by aiming to constantly measure, intervene and modify behaviors (Antón, 2012). In this respect, teachers plan specific challenging tasks accompanied by scaffolding (e.g., guidance, cues and modelling) to enhance the ZPD level. This allows students to move from their current abilities to higher levels of competence (Anwar et al., 2024). Consequently, the emphasis in teaching shifts to teachers' potential to implement differentiated instruction tailored to individual aptitudes (Wang & Feng, 2023). Significantly, the ZPD framework facilitates the planning of remedial work that meets all students' levels, specifically in mixed-ability classes. Through systematic analysis of test scores, teachers can identify students' ZPD and improve test item quality.

***Guttman chart analysis of language test scores***

As stated previously, the Guttman chart can be utilized to transform test scores into a readily interpretable visual form, allowing for the interpretation of the test scores to inform teachers' instructional decision-making. Griffin et al. (2014) view the Guttman chart as "a combination of approaches for analyzing assessment data, allowing teachers to gain a picture of their class's current learning and their future learning needs from assessment data" (p. 191). Drawing on this, scores from objective tests can undergo systematic analysis to identify students' areas of strength and weakness for planning corresponding remedial work. In this case, the Guttman chart provides a visual representation of the zone in which a student's performance breaks down from mostly correct answers to mostly incorrect answers, hence, determining students' ZPD. Concerning this, Maggino (2024) argues that Guttman scaling stresses a unidimensional cumulative scaling technique that orders test items hierarchically from easiest to the most difficult. This permits the identification of progression in student performance and interprets which items indicate mastery of the target components.

Based on teachers' interpretation, students can be grouped according to their similar levels. Besides facilitating curriculum design and lesson planning within the framework of the ZPD (Anwar et al., 2024), Guttman chart analysis also allows teachers to revise test items at both general and item-specific levels to enhance their validity and reliability. As discussed previously, in EFL classrooms, test results should serve formative purposes, while acting as reliable tools that drive a wide range of high-stakes decisions, such as accountability and placement. For this reason, rigorous analysis without sophisticated statistical knowledge becomes critical to ensure that decisions are based on valid and reliable data. Relevant to this, Ismail and Yoestara (2017) found that EFL teachers rarely examine the quality of test items when devising a language test, highlighting the pressing need for a practical and systematic method for evaluating the language test items' validity and reliability.

***Procedural steps for devising a Guttman chart***

To create a Guttman chart, it is necessary to follow the steps below sequentially. These steps are based on the researcher's experimentation, as no study has yet applied this procedure to devise a Guttman chart to analyze test scores. Initially, the examiner enters students' test scores into an Excel spreadsheet or Google sheet, with the first column containing the test takers' identifiers (e.g., names or class numbers) and the first row including the test items, preceded by their corresponding scoring weights. Next, the examiner enters each test taker's item scores as microdata at the item level under its specific item category and scoring weight, ideally, with '0' indicating an incorrect answer and '1' indicating a correct answer.

Once the test microdata are entered for all students, the examiner sums each student's row. This gives the total test score a student received in all the correct items. Then, the examiner sums each column and divides it by its corresponding scoring weight, ideally, '1'. This indicates the standardized score of each item and the total number of students who responded to the given item correctly. To facilitate this process, only the first column is calculated, then the selection box is left-clicked and dragged to the right across the blank cells for the rest of the items. This will automatically calculate the standardized scores for the rest of the items. The same process applies to tallying the rest of the students' total test scores.

Having done so, it is time to order the students' total test scores in ascending order. To do this, the examiner selects the first student tally from the top right-hand corner, then drags the cursor diagnostically to the left to select all students' identifiers (e.g., names or class numbers), including all of their scores while excluding the item labels and their scoring weights. Next, on the toolbar, the examiner chooses 'sort and filter'. In the 'sort by' instruction, the examiner selects the assigned label 'total score' as the option for sorting. Then, in the 'sort on', the examiner selects 'smallest to largest' before pressing 'Okay'. This will complete the sorting of the Guttman chart in ascending order based on students' total scores. Lastly, the standardized scores of the items are sorted in descending order. The examiner could do this by repeatedly inserting empty columns and copying and pasting each column in a descending order based on its corresponding standardized score.

In the end, the Guttman chart will display students' total test scores, with the lowest at the top and the highest at the bottom, and the items' standardized scores ordered in descending order from left to right. Having done so, the last step for the examiner is to select all of the test item scores (i.e., the 1's and 0's), then click on conditional formatting, located on the toolbar. Having added conditional formatting, the visual representation of students' test item scores will be displayed in a color-coded format, indicating anomalies in the data in a different color. Drawing on the examiner's interpretation skills, along with the class profile knowledge, this will suggest the items needing remedial work or targeted teaching, in addition to potential individual items needing revision.

## Method

This section presents the research design, participants, ethical considerations, data collection instruments, including validity and reliability issues, and analysis procedures.

### *Research design*

This study adopts a quantitative approach-postpositivist worldview experimental strategy (Creswell, 2009), with a repeated-measures pre-test and post-test design. The overall aim of this study is to examine student participants' change in post-test performance as a result of a remedial work (Mills & Gay, 2016). The remedial work was prepared based on the Guttman chart analysis of the objective written pre-test scores to identify the participants' ZPD.

### *Participants*

The participants in this study were 25 high school students studying in a common core science class at Abdelhadi Boutaleb High School, Sale city, Morocco, who attend three sessions per week of EFL as a supplementary school subject. They were between 15 and 16 years old. Concerning their gender, 11 participants were males, and 14 were females.

### *Ethical considerations*

Participation in this study was voluntary. In addition, students were informed in advance about the objective of the study and the estimated completion time. Approval was obtained from the school administration in advance. Anonymity and confidentiality of participants' data were ensured by removing any sensitive personal information, such as names.

### *Instruments*

This study employed two objective tests to measure the participants' achievement in EFL knowledge of vocabulary, grammar and communication sampled from the common core science syllabus, particularly, units 5 and 6 of the textbook entitled: "*Visa to the World*". Below is a description of the written pre-test, the remedial work and the online post-test. These instruments were developed by the researcher following the local assessment guidelines (e.g., NCEEO, 2007).

***Pre-test***

The pre-test consists of three sections: vocabulary, grammar and communication. The vocabulary section focuses on lexis related to the units of "Health" and "Leisure" while the grammar section encompasses various structures such as comparatives and superlatives, modals (can, must and mustn't), future tenses (i.e., will and be going to), present simple and continuous tenses, in addition to conjunctions. Besides, the communication section includes communicative functions, such as opinions, agreements, invitations, and permission. The test uses mainly SR items, with tasks progressively increasing in complexity within each section. The test is scored dichotomously (0 or 1) using a norm-referenced scale of 20/20. The subsequent sections outline the aims, the items and the scoring weights for each component.

*Vocabulary*

This section consists of three tasks. First, the two-item chart-filling task prompts students to recall the specific health problems corresponding to the given body parts. Second, a gap-fill/list task presents three prompt items with five choices each, measuring students' recognition of vocabulary items related to "Leisure". Third, a categorizing task with one item is provided to recall vocabulary related to "Leisure". Table 1 provides further details about this section:

Table 1. The pre-test vocabulary section

| Theme | Description | | | |
|---|---|---|---|---|
| | Aim | Tasks | Number of Items | Weight |
| Health | Recall | Chart Completion | 2 | 2 pts (1/2) |
| Leisure | Recognition | Gap-fill/list | 3 | 3 pts (1/3) |
| Leisure | Recall | Categorizing | 1 | 1 pt |
| Total | | 3 | 6 | 6 pts |

*Grammar*

This section encompasses six tasks. Initially, a one-item multiple choice task assesses students' recognition of a short form of comparative, besides a two-item multiple choice task assessing students' recognition of modal verbs (must, mustn't, can). In addition, a one-item word formation task evaluates students' accurate use of the superlative form. Besides, a two-item sentence transformation task measures students' recall of future tense forms (i.e., will and going to). Further, a tense-shift task comprising two items assesses students' use of present simple and continuous forms. Finally, a gap-fill task with a list measures students' use of conjunctions (e.g., and, but, so, etc.). Table 2 describes this section:

Table 2. The pre-test grammar section

| Structures | Description | | | |
|---|---|---|---|---|
| | Aim | Tasks | Number of Items | Weight |
| Comparative (short form) | Recall | M. choices | 1 | 1 pt |
| Modals (can, must, mustn't) | Recall | M. choices | 2 | 2 pts (1/2) |
| Superlative (long form) | Use | Word form | 1 | 1 pt |
| Future with 'will' & 'be going to' | Recall | Sentence. T | 2 | 2 pts (1/2) |

| Structures | Description | | | |
|---|---|---|---|---|
| | Aim | Tasks | Number of Items | Weight |
| Present simple/ Continuous | Use | Tense shift | 2 | 2 pts (1/2) |
| Conjunctions | Use | Gap-fill/ list | 2 | 2 pts (1/2) |
| Total | | 6 | 10 | 10 pts |

*Communication*

This section consists of two tasks, a matching task with two items measuring students' ability to recall communicative functions. It also includes two tasks of responding to situations to assess students' appropriate use of these communicative acts. Table 3 describes this section:

Table 3. The pre-test communication section

| Com. Functions | Description | | | |
|---|---|---|---|---|
| | Aim | Tasks | Number of Items | Weight |
| Appropriate Response | Recall | Matching | 2 | 2 pts (1/2) |
| Exchange Completion | Use | R. to situations | 2 | 2 pts (1/2) |
| Total | | 2 | 4 | 4 pts |

***Remedial work plan***

The remedial work consists of three sections, namely vocabulary, grammar and communication. Each section emphasizes the tasks that match the identified ZPD of the students. To elaborate, based on the Guttman chart analysis, the majority of students did not respond to the following tasks respectively: responding to situations (communicative functions), chart completion (Health vocabulary), word formation (superlative/comparative), matching (communicative functions), sentence transformation (future forms) and tense shift (present simple/continuous). Accordingly, the examiner designed a worksheet targeting these constructions using various SR and CR tasks to remedy students' weaknesses. Table 4 describes each section separately.

Table 4. The sections of the remedial work

| Theme | Description | | |
|---|---|---|---|
| | Aim | Tasks | Number of Items |
| Health | Recall | Matching | 6 |
| Health and Leisure | Recall | Gap-fill/list | 5 |
| Leisure | Recall | Categorizing | 2/8 choices |
| Total | | 3 | 13 |
| Structures | Description | | |
| | Aim | Tasks | Number of Items |
| Tenses (present /future) | Use | Sent. Transformation | 4 |
| Present (simple/continuous) | Use | Tense shift | 7 |
| Comparative/Superlative | Use | Word formation | 6 |
| Total | | 3 | 17 |
| Com. Functions | Description | | |
| | Aim | Tasks | Number of Items |
| Appropriate Response | Use | R. to situations | 6 |

| Theme | Description | | |
|---|---|---|---|
| | Aim | Tasks | Number of Items |
| Writing & Performing Exchanges | Use | Roleplay | 4 |
| Total | | 2 | 10 |

***Post-test***

The post-test involves three sections: vocabulary, grammar and communication sampling content, structures and communicative functions based on Guttman chart analysis of the test scores as well as students' observed performance during the remedial work session. The following account briefly describes each language component.

*Vocabulary*

The vocabulary section measures students' recall of vocabulary related to themes of "Health and Leisure" using multiple-choice questions. Table 5 describes this section:

Table 5. The post-test vocabulary section

| Theme | Description | | | |
|---|---|---|---|---|
| | Aim | Tasks | Number of Items | Weight |
| Leisure | Recall | M. Choices | 3 | 3 pts (1/3) |
| Health | Recall | M. Choices | 2 | 2 pts (1/2) |
| Total | | 2 | 5 | 5 pts |

*Grammar*

The grammar section includes multiple choice questions targeting the following constructions: present simple and continuous, comparative and superlative forms and modal verbs. Table 6 describes this section:

Table 6. The post-test grammar section

| Structures | Description | | | |
|---|---|---|---|---|
| | Aim | Tasks | Number of Items | Weight |
| Present (simple/continuous) | Recall | M. Choices | 4 | 4 pts (1/4) |
| Comparative/ Superlative | Recall | M. Choices | 4 | 4 pts (1/4) |
| Modal verbs | Recall | M. Choices | 2 | 2 pts (1/2) |
| Total | | 3 | 10 | 10 |

*Communication*

The communication section measures primarily students' recall of the communicative functions, namely expressing necessity, agreeing, asking for permission and making and responding to suggestions. Table 7 provides extra details about this section:

Table 7. The post-test communication section

| Com. Functions | Description | | | |
|---|---|---|---|---|
| | Aim | Tasks | Number of Items | Weight |
| Appropriate Response | Recall | M. Choices | 4 | 4 pts (1/4) |
| Exchange Completion | Use | R. to situations | 1 | 1 pt |
| Total | | 2 | 5 | 5 |

As an overall observation, the pre-test was administered in written form. It included mainly various SR items (i.e., multiple-choice and matching), besides CR items (i.e., gap-fill/list, chart completion, sentence transformation, tense shift and word formation). In contrast, the online post-test primarily utilized multiple-choice items. The choice of online testing was due to its timing coinciding with the closure of the second semester. Additionally, the test format was limited to multiple-choice items to accommodate students' unfamiliarity with online testing.

***Validity and reliability concerns***

Content validity was accounted for by ensuring alignment between test items and the learning objectives of the units targeted. Moreover, the tests and the remedial work were shared with two experienced EFL teachers for feedback on the clarity of the instructions and the representativeness of the items in relation to the intended constructs. As for reliability, most items of both tests were SR items, ensuring objectivity in scoring. Further, despite the difference in the delivery mode, identical administration conditions were maintained for the pre-test and post-tests, including completion time and testing environment.

***Data collection procedure***

Data were collected using pre-test and post-test measures. A written pre-test was first administered to a Common Core science class taking EFL sessions. The scores obtained from this test were analysed using a Guttman chart to identify students' ZPD. This analysis also facilitated the identification of item-related sources of anomalies. Based on these results, a remedial work was designed to scaffold students and address their areas of weakness. Subsequently, a post-test was administered to measure the effectiveness of the remedial work. Figure 2 summarizes the data collection process.

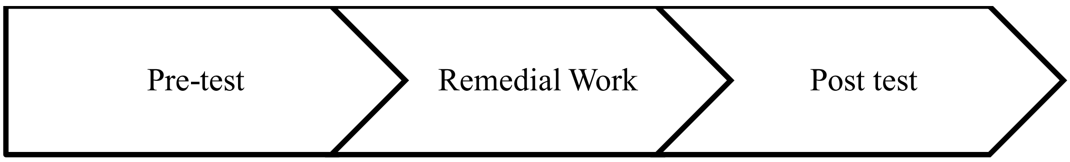


Figure 2. The data collection process

***Data analysis procedure***

The scores of the written pre-test were subjected to Guttman chart analysis to primarily identify students' ZPD, including potential sources of anomalies in the test items. Based on this analysis, an immediate remedial work intervention was devised to remedy the students' weaknesses. Subsequently, the scores of the objective online test were also analyzed using the Guttman chart. This allowed a comparison of students' progress from the pre-test to the post-test and an evaluation of the functionality of the online test items.

Furthermore, students' scores from the two tests were analyzed using SPSS 26.0 (IBM Corporation, Armonk, NY) to compare the differences between the pre-test and post-test scores due to the effect of the remedial work intervention. Since the data violated the normality assumption, the non-parametric Wilcoxon Matched-Pairs Signed Ranks Test (see Table 8) was employed as it considers both the degree of change and the direction of the difference (Hatch & Lazaraton, 1991). In concrete terms, this was meant to compare the differences between paired pre-test and post-test scores. Following this, eta squared ($\eta^2$) (see Figure 3) was calculated to determine the proportion of variance in the ranks attributable to the remedial intervention.

$$\eta^2 = \frac{z^2}{N - 1}$$

Figure 3. The formula for calculating eta squared ($\eta^2$)

## Findings and Discussion

### *Findings*

This section presents the Guttman chart analysis of the written pre-test scores, aiming to identify students' ZPD, design a remedial work, and evaluate the functionality of the test items. Besides, it provides the Guttman chart analysis of the post-test scores, intended primarily to compare students' performance on both tests and assess the items of the online post-test. Furthermore, it highlights the statistical analysis comparing pre-test and post-test scores.

*Guttman chart analysis of the pre-test scores*

| S.W | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S.N | V GF I1 | G MC I1 | G GF I1 | V GF I2 | V GF I3 | C M I1 | V CC I1 | G MC I1 | G MC I2 | G TS I1 | V CT I1 | G GF I2 | G TS I2 | G ST I1 | G ST I2 | C M I2 | G WF I1 | V CC I2 | C RS I1 | C RS I2 | T |
| 19 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| 8 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 7 |
| 9 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 7 |
| 10 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 8 |
| 14 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 |
| 16 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 8 |
| 2 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 8 |
| 4 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 8 |
| 23 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 8 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 9 |
| 28 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 9 |
| 18 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 10 |
| 29 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 10 |
| 6 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 11 |
| 30 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 11 |
| 17 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 12 |
| 26 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 12 |
| 12 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 14 |
| 20 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 14 |
| 21 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 14 |
| 22 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 15 |
| 24 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 17 |
| 11 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 18 |
| 31 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 19 |
| 25 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 19 |
| S.S | 24 | 24 | 24 | 20 | 18 | 16 | 16 | 15 | 15 | 15 | 14 | 13 | 12 | 12 | 10 | 8 | 8 | 8 | 7 | 3 | |

*Note*: S.W: Scoring Weight, S.N: Student Number, S.S: Standardized Score, T: Total, I: Item V: Vocabulary, G: Grammar, C: Communication, GF: Gap Fill, MC: Multiple Choice, M: Matching, TS: Tense Shift, ST: Sentence Transformation, WF: Word Formation, RS: Responding to Situations, CC: Chart Completion, CT: Categorizing.

Figure 4. Guttman chart analysis of the pre-test scores

As shown in Figure 4, Guttman chart analysis of the written pre-test demonstrates students' language test scores on different items. It shows the total test score of the correct items each student received. It also displays the standardized scores, indicating the total number of items answered more or less by students. These item scores are color-coded, with green indicating lower performance and white indicating higher performance.

Top performers are students whose class numbers are 22, 24, 11, 31, and 25 while 11 students got below average 10/20. At the item level, gap-fill (I1, Vocabulary), multiple choice (I1, Grammar), gap-fill (I1, Grammar), and gap-fill (I2, Vocabulary) seem to have the highest scores. Conversely, the majority of the items, namely responding to situations (I2 and I1, Communication), chart completion (I2, Vocabulary), word formation (I1, Grammar), matching (I2, Communication), sentence transformation (I1 and I2, Grammar) and tense shift (I2, Grammar) demonstrate respectively lower scores, suggesting the ZPD of the class, the areas needing remedial work.

*Guttman analysis of the post-test scores*

| S.W | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 20 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S.N | V MC I3 | V MC I4 | V MC I2 | G MC I8 | G MC I10 | C MC I3 | G MC I4 | G MC I7 | G MC I6 | C MC I1 | V MC I1 | G MC I2 | G MC I1 | G MC I9 | V MC I5 | G MC I3 | C MC I4 | C RS I5 | G MC I5 | C MC I2 | T |
| 16 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |
| 2 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| 9 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 8 |
| 19 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 | 9 |
| 4 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 12 |
| 10 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 12 |
| 8 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 13 |
| 29 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 13 |
| 18 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| 30 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 | 13 |
| 12 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 15 |
| 17 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 16 |
| 26 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 16 |
| 5 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 17 |
| 11 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 17 |
| 20 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 18 |
| 22 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 18 |
| 6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 18 |
| 14 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 18 |
| 23 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 18 |
| 21 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 19 |
| 25 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 19 |
| 28 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 19 |
| 24 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 19 |
| 31 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 20 |
| S.S | 25 | 23 | 22 | 22 | 22 | 22 | 21 | 21 | 21 | 21 | 20 | 19 | 18 | 17 | 16 | 15 | 15 | 12 | 11 | 8 | |

*Note*: S.W: Scoring Weight, S.N: Student Number, S.S: Standardized Score, T: Total, I: Item, V: Vocabulary, G: Grammar, C: Communication, MC: Multiple Choice, RS: Responding to Situations

Figure 5. Guttman chart analysis of the post-test scores

Figure 5 demonstrates the Guttman chart analysis of the post-test. Based on students' total scores, the standardized scores and the color-coded scores, it is clear that only 4 students scored below average while about 15 students scored 15 out of 20 or above.

Top performers are students whose class numbers are 31, 24, 28, 25, and 21, respectively. Besides, the Guttman chart analysis also demonstrates that almost all high-performing students answered nearly all items correctly, with some exceptions. However, the majority of students did not answer the following five items respectively: multiple-choice (I2, Communication), responding to situations (I5, Communication), multiple choice (I4, Communication) and multiple choice (I3, Grammar), indicating the need to question the functionality of these items and assess whether students may require further revision of the content they target.

*Wilcoxon matched-pairs signed-rank test*

Table 8. Mean rank of the pre-and post-tests

| Ranks | | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| Post-test - Pre-test | Negative Ranks | 3[a] | 6.17 | 18.50 |
| | Positive Ranks | 21[b] | 13.40 | 281.50 |
| | Ties | 1[c] | | |
| | Total | 25 | | |

a. Post-test < Pre-test
b. Post-test > Pre-test
c. Post-test = Pre-test

Table 8 presents the mean and total ranks for the pre-test and post-test scores. Among the 25 cases, 3 (negative ranks) indicate a decrease in post-test scores compared to pre-test scores, while 21 (positive ranks) show an increase. Only 1 case reflects no change between pre-test and post-test scores. Overall, the data highlights that the majority of students (n = 21) showed improvement, with a few experiencing a decline (n = 3) or no change (n = 1).

Table 9. Test statistics

| | Post-test - Pre-test |
|---|---|
| Z | -3.768[b] |
| Asymp. Sig. (2-tailed) | .000 |

a. Wilcoxon Signed Ranks Test
b. Based on negative ranks.

A Wilcoxon Matched-Pairs Signed-Ranks Test demonstrated a significant difference in the pre-test ($M$ = 11.28, $SD$ = 3.96) and the post-test ($M$ = 14.84, $SD$ = 4.29) scores, Z = -3.768, $p < .001$, indicating that the remedial work led to a statistically significant improvement in test scores. Therefore, the null hypothesis for the third research question, which posits no difference between pre-test and post-test scores due to the remedial work, is rejected.

*Calculating eta squared*

Having applied the Eta Squared ($\eta^2$) formula (see Figure 3) to determine the proportion of variance in the ranks accounted for by the remedial work, the computed value is 0.59, indicating that 59% of the variance in the ranks can be attributed to the effect of the remedial work, which was designed using Guttman chart analysis.

***Discussion***

Due to their practicality (Giraldo, 2020), objective tests are frequently used in Moroccan EFL classrooms (Benzehaf, 2017; El zerk, 2025a; El zerk & Ichebah, 2025; Mamad & Vigh, 2021). This study demonstrated that these objective tests can play a crucial formative role when the test scores are analyzed to identify students' current developmental stages within the framework of ZPD (Anwar et al., 2024; Vygotsky, 1978; Wang & Feng, 2023), thereby facilitating remedial work intended for differentiation and scaffolding (Nouri, 2021). This supports the view that language tests that meet statistical requirements and accurately represent language constructs (Fox, 2016; Giraldo, 2020) can offer teachers valuable information about specific items (Bailey, 2016). Teachers can then interpret these data to make formative decisions, including multi-level groupings, planning of instructional materials (Nouri, 2021), designing instruction, and curriculum reforming (Bonner, 2013; Fan et al., 2021), by analyzing the gap between the present and the expected performance (Schneider et al., 2013).

The results of this study demonstrated a significant improvement in students' scores from the pre-test to the post-test as a result of the remedial work developed based on Guttman chart analysis. To elaborate, the Guttman chart analysis of the pre-test scores revealed an immediate need for remedial work, with four students scoring the highest and eleven scoring below average. Furthermore, the chart graphically identified the specific items that most students found challenging, indicating the need for differentiated instruction and scaffolding tailored to students' needs (Anwar et al., 2024; Wang & Feng, 2023). It also highlighted items that were sources of anomalies needing review. Consequently, the Guttman chart enables effective analysis of students' test scores and supports the improvement of teacher-made tests (Ghaicha, 2016) by enabling practical meta-analysis of test items for future use (Ismail & Yoestara, 2017). More importantly, this study addressed the gap in understanding how teachers use assessment data to inform teaching (Jang, 2017), especially with the absence of explicit guidelines for the formative use of assessment data (van der Steen et al., 2022).

The Guttman chart analysis of the post-test scores further supports that the observed improvement is likely due to the remedial work. The latter relatively increased all individual students' test performance, except for three (students 16, 2 and 9) whose test scores from the pre-test and post-test were (8 vs. 5), (8 vs. 6) and (7 vs. 8), respectively. This corroborates the view that objective tests do not meaningfully explain the reasons behind students' performance (Russel, 2020), highlighting the need to frame teachers' interpretation of assessment results within students' current levels of understanding and their profiles of strengths and weaknesses (Kane & Wools, 2019). Therefore, it is crucial to interpret these scores in conjunction with the teachers' contextual knowledge of students' in-class and out-of-class performance to identify the gap between current and expected performance (Schneider et al., 2013). Furthermore, while objective tests often provide accurate measures of students' competence that meet psychometric criteria (Fox, 2016; Giraldo, 2020), they typically require additional data for a comprehensive assessment, as they fail to explain the reasons behind students' performance (Russell, 2019).

The Guttman chart analysis of the pre-test highlighted, through color-coded data, the need to reassess the sources of anomalies in the functionality of various items, as well as students' understanding of the target constructs. However, the limited number of items targeting each construct compromises the validity of the assumptions underlying the preparation of the post-test. This implies the importance of measuring a limited number of constructs with specific and adequate objective items, sequenced in terms of complexity. In this respect, Guttman scaling of items emphasizes a unidimensional cumulative scaling hierarchy from easiest to the most difficult (Maggino, 2024). This approach ensures that the test information can effectively guide future instructional planning by providing reliable inferences (Jang, 2017), especially in high-stakes testing contexts, such as Morocco, where tests and quizzes are pervasive (Benzehaf, 2017; El zerk, 2025a; El zerk & Ichebah, 2025; Mamad & Vigh, 2021).

In the post-test, Guttman chart analysis revealed a statistically significant improvement in item functionality. High-performing students demonstrated improved performance by answering almost all items. Conversely, only a few lower-performing students, specifically students numbered 16, 2, 9, and 19, missed more higher-functioning items. This highlights both improved reliability and the need for further remedial work. In contrast to the written pre-test, which included various item formats, the online post-test predominantly featured multiple-choice items. This shift raises considerations about the validity of its scores. This corroborates Ismail and Yoestara's (2017) findings that EFL teachers take limited precautions to ensure test item quality. Nevertheless, online tests proved to be more practical for collecting objective test scores. Besides, high-quality multiple-choice items are particularly featured as valid and reliable. However, it is worth emphasizing that tests provide initial diagnostic inferences which require incorporating other criterion-referenced and performance-based assessment techniques. This is essential to account for the issue of indirect underachievement (Nouri, 2021). Importantly, triangulating diverse data collection will integrate assessment with teaching to reveal emerging abilities (Poehner et al., 2017).

## Conclusion

Based on the results of this study, the Guttman chart facilitated the analysis of students' performance in an objective, discrete-point EFL test to identify their ZPD. Consequently, it is a practical method for informing future targeted teaching or the planning of remedial work, alongside identifying sources of anomalies; thereby, it enhances the validity and reliability of the test item bank. However, it is worth highlighting that the analysis and interpretation of test scores is insufficient if not accompanied by other criterion-referenced assessment data to draw valid conclusions about students' weaknesses.

This study demonstrated that Guttman chart analysis is more practical and feasible for analyzing online test scores. Yet, the insufficient items measuring each construct and the lack of consistency between the pre-test and post-test in terms of difficulty, content and format may have affected the validity of the observed changes occurring between the pre-and post-test scores as a result of the remedial work. Therefore, it is recommended to replicate this study within the EFL context using online objective tests created using Google Forms. This allows for ready-made data downloads instead of manual entry. Furthermore, using Guttman chart analysis for piloting standardized high-stakes test items is recommended to enhance their reliability and validity.

Based on the results of this study, Guttman chart analysis is effective only when objective test items are scored dichotomously (1 or 0). It also requires limiting the sampling domain to specific skills and knowledge areas, allocating a sufficient number of items per content domain, and arranging them in terms of complexity. Further, multiple-choice items are highlighted as an effective measurement task that produces practical and consistent scores, serving as primary diagnostic assessment data as part of a holistic and triangulated assessment.

While this study was the first endeavor to employ the Guttman chart for objective test analysis, it is critical to acknowledge some of its limitations. A key limitation is that it relied exclusively on the quantitative data, drawing on the Guttman chart analysis findings, along with pre-test and post-test findings. Future research may incorporate qualitative data collection instruments for deeper insights from the lens of students on the effectiveness of the Guttman chart in analyzing test scores for developing a corresponding remedial work. Another limitation worth highlighting is the scarcity of theoretical and empirical foundations of the Guttman chart in language assessment research, which constrains the generalizability of this study's findings. Furthermore, differences in delivery mode and sampling between the pre-test and post-test may constitute a source of unintended bias.

Future research should examine whether SR items (e.g., multiple-choice) and CR items (e.g., responding to situations) equally measure depth of knowledge and skill level. In addition, it should examine whether these item formats interact with other student characteristics, such as gender, major, subject matter, or school level. Moving forward, true experimental designs are needed to validate the effectiveness of Guttman chart analysis in diagnosing students' ZPD and identifying the sources of errors, particularly considering the limitations of this study (e.g., delivery mode, target constructs, and sufficient objective items). Besides, it is worth investigating whether frequent use of objective tests is associated with increased language proficiency, including improved performance on high-stakes tests. Furthermore, in the Moroccan EFL context, it is worth considering the integration of online assessments with the development of software that enables learning analytics.

**References**

Antón, M. (2012). Dynamic assessment. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 106-119). Routledge. https://doi.org/10.4324/9780203181287

Anwar, M. N., Mushtaq, N., Mubeen, A., & Iqbal, M. (2024). The power of ZPD: Enhancing teaching and learning. *Journal of Education and Social Studies, 5*(2), 396-405. https://doi.org/10.52223/jess.2024.5220

Bachman, L. F. (2004). *Statistical analyses for language assessment book.* Cambridge University Press.

Bailey, A. L. (2016). Assessing the language of young learners. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (pp. 1-20). Springer. https://doi.org/10.1007/978-3-319-02326-7_22-1

Benzehaf, B. (2017). Exploring teachers' assessment practices and skills. *International Journal of Assessment Tools in Education, 4*(1), 1-18. https://doi.org/10.21449/ijate.254581

Bonner, S. (2013). Validity in classroom assessment: Purposes, properties, and principles. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 87-106). SAGE Publications. https://doi.org/10.4135/9781452218649.n1

Brookhart, S. M. (2019). Feedback and measurement. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 63-78). Routledge. https://doi.org/10.4324/9780429507533-5

Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches* (3rd ed.). SAGE Publications.

El zerk, M. (2025a). Classroom-based assessment techniques for EFL skills and components in Moroccan high schools. *International Jordanian Journal Aryam for Humanities and Social Sciences (IJJA), 7*(2), 105-122. https://doi.org/10.5281/zenodo.16606159

El zerk, M. (2025b). Classroom-based assessment purposes from the lens of Moroccan EFL high school teachers. *Studies in Humanities and Education*, *6*(1), 21-34. https://doi.org/10.48185/she.v6i1.1549

El zerk, M. & Ichebah, M. (2025). Classroom-based assessment practices and assessment literacy in Moroccan EFL high schools. *Journal of Applied Language and Culture Studies*, *8*(2), 198-218.

Fan, T., Song, J., & Guan, Z. (2021). Integrating diagnostic assessment into curriculum: A theoretical framework and teaching practices. *Language Testing in Asia, 11*, Article 2. https://doi.org/10.1186/s40468-020-00117-y

Ferrara, S., Maxey-Moore, K., & Brookhart, S. M. (2019). Guidance in the standards for classroom assessment: Useful or irrelevant? In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 97-119). Routledge. https://doi.org/10.4324/9780429507533-7

Fox, J. (2016). Using portfolios for assessment/alternative assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (3rd ed., pp. 1-13). Springer. https://doi.org/10.1007/978-3-319-02261-1_26

Giraldo, F. D. (2020). Validity and classroom language testing: A practical approach. *Colombian Applied Linguistics Journal, 22*(2), 194-206. https://doi.org/10.14483/22487085.15998

Ghaicha, A. (2016). Theoretical framework for educational assessment: A synoptic review. *Journal of Education and Practice, 7*(24), 212-231.

Griffin, P., Robertson, P., & Hutchinson, D. (2014). Modified Guttman analysis. In P. Griffin (Ed.), *Assessment for teaching* (pp. 187-211). Cambridge University Press.

Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* Newbury House Publishers.

Ismail, N. M., & Yoestara, M. (2017). Teachers' behalf on language test construction. *LLT Journal: A Journal on Language and Language Teaching, 20*(2), 78-84. https://doi.org/10.24071/llt.v20i2.734

Jang, E. E. (2017). Cognitive aspects of language assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (pp. 163-177). Springer. https://doi.org/10.1007/978-3-319-02261-1_11

Kane, M. T., & Wools, S. (2019). Perspectives on the validity of classroom assessments. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 11-26). Routledge. https://doi.org/10.4324/9780429507533-2

Leighton, J. P. (2019). Cognitive diagnosis is not enough: The challenge of measuring learning with classroom assessments. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 27-45). Routledge. https://doi.org/10.4324/9780429507533-3

Maggino, F. (2024). Guttman scale. In F. Maggino (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 2877-2881). Springer. https://doi.org/10.1007/978-3-031-17299-1_1218

Mamad, A., & Vígh, T. (2021). Moroccan EFL public high school teachers' perceptions and self-reported practices of assessment. *Journal of Language and Education, 7*(3), 119-135. https://doi.org/10.17323/jle.2021.12067

McMillan, J. H. (2019). Discussion of Part I: Assessment information in context. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 79-94). Routledge. https://doi.org/10.4324/9780429507533

Mills, G. E., & Gay, L. R. (2016). *Educational research: Competencies for analysis and applications.* Pearson Education. https://doi.org/10.37074/jalt.2018.1.2.14

National Centre for Evaluation, Exams and Orientation (NCEEO). (2007). *Secondary qualifying educational assessment of a foreign language: English, Spanish, Italian, and German (Ministerial Circular* No. 142-07*).* National Centre for Evaluation, Exams and Orientation. https://eflcollective.com/Other/teacher-materials/ministerial-notes/

Nouri, N. (2021). Formative testing program for undergraduate EFL Moroccan students. *European Journal of Education Studies, 8*(4), 317-326. https://doi.org/10.46827/ejes.v8i4.3699

Poehner, M. E., Davin, K. J., & Lantolf, J. P. (2017). Dynamic assessment. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (pp. 243-256). Springer. https://doi.org/10.1007/978-3-319-02261-1_18

Randel, B., & Clark, T. (2013). Measuring classroom assessment practices. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 145-163). SAGE Publications. https://doi.org/10.4135/9781452218649.n1

Rodriguez, M. C., & Haladyna, T. M. (2013). Writing selected-response items for classroom assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 293-312). SAGE Publications. https://doi.org/10.4135/9781452218649.n17

Russell, M. (2019). Digital technologies: Supporting and advancing assessment practices in the classroom. In S. M. Brookhart & J. H. McMillan (Eds.), *Classroom assessment and educational measurement* (pp. 224-242). Routledge. https://doi.org/10.4324/9780429507533-13

Settiawana, D., & Hilmawan, R. (2016). Increasing transparency in assessment to improve students' learning at the Language Development Centre of UIN Suska Riau. In *Proceedings of the International Conference on Teacher Training and Education* (Vol. 1, No. 1, pp. 365-373).

Schneider, M. C., Egan, K. L., & Julian, M. W. (2013). Classroom assessment in the context of high-stakes testing. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 55-70). SAGE Publications. https://doi.org/10.4135/9781452218649.n4

Tsagari, D., & Cheng, L. (2017). Washback, impact, and consequences revisited. In E. Shohamy, I. Or, & S. May (Eds.), *Language testing and assessment. Encyclopedia of language and education* (3rd ed., pp. 359-372). Springer. https://doi.org/10.1007/978-3-319-02261-1_24

Tomlinson, C. A., & Moon, T. R. (2013). Differentiation and classroom assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 415-430). SAGE Publications. https://doi.org/10.4135/9781452218649.n23

Turner, C. E. (2012). Classroom assessment. In G. Fulcher & F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 65-78). Routledge. https://doi.org/10.4324/9780203181287

van der Steen, J., van Schilt-Mol, T., van der Vleuten, C., & Joosten-ten Brinke, D. (2022). Supporting teachers in improving formative decision-making: Design principles for formative assessment plans. *Frontiers in Education, 7*, 182-194. https://doi.org/10.3389/feduc.2022.925352

Vygotsky, L. S. (1978). Interaction between learning and development. In M. Cole, V. Jolm-Steiner, S. Scribner, & E. Souberman (Eds.), *Mind in society: Development of higher psychological processes* (pp. 79-91). Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4.11

Wang, Z., & Feng, J. (2023). Teaching quality evaluation based on student's Zone of Proximal Development (ZPD) prediction. *Journal of Education and Educational Research, 6*(1), 44-48. https://doi.org/10.54097/jeer.v6i1.14157

Widyastuti, P., & Harto Pramono, Y. G. (2021). An analysis of English teachers' written feedback in primary students' report cards. *LLT Journal: A Journal on Language and Language Teaching, 24*(2), 529-556. https://doi.org/10.24071/llt.v24i2.2880

Xie, Q., & Jia, Q. (2025). Three decades of research on washback (1993–2023): A bibliometric study. *Language Testing in Asia, 15*(23), 1-29. https://doi.org/10.1186/s40468-025-00357-w

Zaretsky, V. K. (2024). Zone of proximal development: Evolution of the concept. *Cultural-Historical Psychology*, *20*(3), 45-57. https://doi.org/10.17759/chp.2024200305