

## A proposed revised version of the scoring rubrics of VSTEP 3-5 level

Phiên bản đề xuất chỉnh sửa bộ tiêu chí đánh giá kỳ thi VSTEP bậc 3-5

Ton Nu My Nhat<sup>a,b\*</sup>  
Tôn Nữ Mỹ Nhật<sup>a,b\*</sup>

<sup>a</sup>*Institute of Linguistics, Duy Tan University, Da Nang, 550000, Vietnam*

<sup>a</sup>*Viện Ngôn ngữ, Đại học Duy Tân, Đà Nẵng, Việt Nam*

<sup>b</sup>*Faculty of English Language, School of Foreign Language, Duy Tan University, Da Nang, 550000, Vietnam*

<sup>b</sup>*Khoa Tiếng Anh, Trường Ngoại Ngữ, Đại học Duy Tân, Đà Nẵng, Việt Nam*

(Ngày nhận bài: 03/11/2020, ngày phân biện xong: 27/02/2021, ngày chấp nhận đăng: 08/03/2021)

### Abstract

VSTEP (Vietnamese Standardized Test of English Proficiency) is a proficiency test in line with the Six-level framework for foreign language proficiency in Vietnam (VNFLPF), equivalent to the CEFR, issued by the MOET dated March 3, 2015. Of the versions for different target groups, the VSTEP 3-5 level is undoubtedly the most wide-spread so far, and common in multiple institutions of higher education in Vietnam. This article is to describe and critically evaluate the issued scoring rubrics of the two productive skills – Speaking and Writing; and basing on an analysis and modification of the descriptors, to propose revised versions of the marking rubrics. The product represents the more concise and clearer rubrics. The revised versions are of practical significance to the practice of English language teaching for graduates and post-graduates, the issue of training the assessors of the two productive skills, and/or the endeavor of self-improvement on the part of instructors of English in order to ensure standardized quality in assessment.

**Keywords:** VSTEP; VSTEP 3-5; scoring rubrics; Six-level framework for foreign language proficiency in Vietnam.

### Tóm tắt

VSTEP (Vietnamese Standardized Test of English Proficiency) là kỳ thi đánh giá năng lực tiếng Anh theo Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam (tương đương với trình độ A1, A2, B1, B2, C1, C2), được Bộ GD&ĐT ban hành vào tháng 03 năm 2015. Trong những phiên bản của VSTEP, dành cho những đối tượng khác nhau, thì bài thi VSTEP bậc 3-5 là phổ biến hơn cả và đang được sử dụng rộng rãi trong các trường đại học ở Việt Nam hiện nay. Bài viết này nhằm mục tiêu miêu tả và đánh giá bộ tiêu chí đánh giá hai kỹ năng Nói và Viết của kỳ thi này, và dựa trên những kết quả phân tích và chỉnh sửa ngôn từ, kết quả nghiên cứu đã cho ra 2 bộ tiêu chí đánh giá ngắn gọn, súc tích, và rõ ràng hơn. Hai bộ tiêu chí đánh giá mới có ý nghĩa đối với thực tiễn dạy-học tiếng Anh bậc đại học, sau đại học, cũng như với thực tế bồi dưỡng đội ngũ cán bộ chấm thi hai kỹ năng Nói, Viết, hay/và thực tế tự bồi dưỡng của giảng viên tiếng Anh để bảo đảm chất lượng trong quá trình chấm thi hướng đến các chuẩn quốc tế.

**Từ khóa:** Kỳ thi đánh giá năng lực tiếng Anh của Việt Nam; Kỳ thi đánh giá năng lực tiếng Anh của Việt Nam bậc 3-5; Tiêu chí chấm thi; Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam.

\*Corresponding Author: Ton Nu My Nhat; Institute of Linguistics, Duy Tan University, Da Nang, 550000, Vietnam; Faculty of English Language, School of Foreign Language, Duy Tan University, Da Nang, 550000, Vietnam  
Email: tnmynhath70@gmail.com

## 1. Introduction

The Vietnamese Standardized Test of English Proficiency (VSTEP) was released according to Decision No. 729/QĐ-BGDĐT, dated March 11, 2015 [1]. It is claimed that this ‘*first-ever made-in-Vietnam standardized EFL tests*’ was developed ‘*intensively with utmost effort and energy*’ [2, p. 74]. There are five VSTEP variants for different targeted groups, of which the VSTEP 3-5 level is the most influential. This multi-level test is aimed to cater for the certification needs of a large number of groups. Many institutions in Vietnam have made use of this specification to award the B1-C1 English proficiency certificates to multiple target groups in recent years. It has widely been used in colleges and examination bodies in Vietnam because it is well developed, reliable, and economical in comparison to the international tests.

Given the wide-spread use of this nationwide test, this article is to argue that the scoring rubrics of the two productive skills of VSTEP 3-5 is far from practical and to propose a revised version of the rubrics. In this article, the currently-circulated scoring rubrics for the two productive skills will first be described in terms of superficial textual features and practicality. The paper will then go on to describe how the rating scales have been revised - the features modified and the procedures, which is followed by the presentation of the revised products. The final section concerns itself with the factors to consider in order to assure scoring quality of these tests.

## 2. The study

### 2.1. A critical analysis of VSTEP 3-5 scoring rubrics

There are three rubrics altogether - one for Speaking test, and two for Writing test - Task 1

and Task 2 [3]. It is claimed that the development of these rating rubrics was based on the Vietnamese six-level language proficiency framework and the other CEFR-aligned rating schemes such as IELTS and Cambridge English exams such as PET, FCE, and CAE.

These analytic scales contain five features for Speaking test - *Grammar, Vocabulary, Pronunciation, Fluency, Organization* - and four features for Writing test - *Task completion, Organization, Vocabulary, Grammar*. All four/five analytical criteria are shared across the whole test and equally weighted. The number of criteria is, thus, within limit [4, p. 193].

Each of the feature has descriptors for 11 points - ascending from 0 to 10. The descriptors are intended to focus the readers’ attention on the test-takers’ strengths, not on their deficiencies. That there are up to 11 points presents a marked weakness. Pollitt (1990, cited in [5, p. 123]) points out that it is ‘*overly optimistic to expect a test of writing to be able to distinguish reliably between five scale points or more.*’ In addition, although the benefits of analytical scoring are to provide valuable diagnostic information for test-takers, only the overall score is reported, which means ‘*a good deal of information provided by the analytic scale is lost*’ [5, p. 120].

It should also be noted that the rules for deriving the overall score are rather complicated: first, each feature is marked on a 11-point scale; then, each score is converted to the assigned level: 4-5.5 equals Level 3; 6-8 equals Level 4; 8.5-10 equals Level 5); finally, the overall score reported is decided, basing on a combination of the four/five individual scores.

A close analysis of the rubrics of two writing tasks reveals that they are different in only the feature of *Task completion*. Another minor

difference can be seen in *Organization*, concerning paragraphing, restricting to one descriptor in the middle scores.

One noticeable weakness of these rubrics lies with the layout. That they are written in Vietnamese and there are too many columns and rows due to the large number of points highly obscures differentiation between the band scores. Lack of meticulous attention to the

layout also results in the ever-highest number of pages of all rubrics we can refer to so far; each rubric takes up five pages. As Luoma [6, p. 81] states, to be useful, the scale descriptors need to be practical, because if they are not brief, it is hard to memorize and use the scale consistently. Table 1 summarizes the superficial textual features of these three scoring rubrics.

Table 1: Textual features of three scoring rubrics

| Features        | SPEAKING scoring rubric | WRITING TASK 1 scoring rubric | WRITING TASK 2 scoring rubric |
|-----------------|-------------------------|-------------------------------|-------------------------------|
| No. of column   | 12                      | 5                             | 5                             |
| No. of row      | 6                       | 12                            | 12                            |
| No. of word     | 2,230                   | 1,259                         | 1,423                         |
| No. of page     | 5                       | 5                             | 5                             |
| No. of features | 5                       | 4                             | 4                             |

To be practical, these weaknesses need taking into consideration and modifications are highly mandated.

## 2.2. Methodology

The procedures were as follows:

(1) The PDF files were converted to Word;

(2) For each feature, the points and the associated descriptors were set in lines, not in columns. For example:

1. *Phát âm tương đối rõ ràng những từ và cụm từ đơn giản nhưng do đôi khi khó nghe nên người đối thoại vẫn phải yêu cầu nhắc lại.*

2. *Phát âm rõ ràng và dễ hiểu những từ và cụm từ, tuy còn mắc khá nhiều lỗi về các âm đơn lẻ và các lỗi về âm khác.*

(3) The number of levels were reduced: Basing on the rules of converting from 11-point scale to 3-5 levels, I combined points 10 and 9 to produce *Level 5*, points 8, 7 and 6 to produce *Level 4*, points 5 and 4 to produce *Level 3*, leaving all the lower points to “*Không xét*” (unqualified for any levels of 3-5).

(4) The descriptors were accordingly grouped into four levels.

(5) Basing on Vietnamese grammar, I conducted a discourse analysis of the group of descriptors in each level and modified the descriptors. Specifically,

- eliminating the repeated descriptors;
- reordering some syntactic elements so that distinctions between the levels would be easier to identify, understand, and remember;
- deleting unnecessary words (mostly modal words, such as ‘chỉ’, ‘được’, ‘một cách’...);
- rewording confusing words and expressions if any.

For example:

Original: *Chỉ sử dụng được một cách hạn chế một số cấu trúc ngữ pháp và mẫu câu đơn giản đã học.*

Revised: *Sử dụng hạn chế một số cấu trúc đơn giản.*

Original: *Đáp ứng đầy đủ và hiệu quả các yêu cầu của đề bài.*

Revised: *Đáp ứng các yêu cầu của đề bài đầy đủ, hiệu quả.*

Original: *Một số lỗi về sự không thống nhất và không phù hợp trong giọng văn.*

Revised: *Một số lỗi về tính thống nhất, phù hợp trong giọng văn.*

Finally, I combined the two rubrics for two tasks of the writing test into only one scale. The only feature which presents to be different, namely *Task completion*, was categorized into the subscales for Task 1 and those for Task 2. Furthermore, the descriptor associated to

paragraphing was maintained as a feature for both tasks because paragraphing also must be observed in this functional genre.

### 3. Findings and discussion

The time-consuming procedures have resulted in the revised versions which bear out many strengths in comparison to the original versions: they are shorter, clearer, more user-friendly, and accordingly more useful and practical (Table 2 and Table 3). Table 4 compares the two versions.

Table 2: Scoring rubrics for speaking test (Tiêu chí chấm điểm bài thi Nói)

| BẬC              | NGŨ PHÁP<br>PHẠM VI<br>ĐỘ CHÍNH XÁC   | TỪ VỰNG<br>PHẠM VI<br>KIỂM SOÁT  | PHÁT ÂM<br>ÂM ĐƠN LẺ<br>TRỌNG ÂM,<br>NGŨ ĐIỀU  | ĐỘ LƯU<br>LOÁT<br>NGẬP<br>NGỪNG<br>NÓI DÀI  | BÓ CỤC<br>NỘI DUNG<br>PHÁT TRIỂN<br>CHỦ ĐỀ<br>MẠCH LẠC<br>LIÊN KẾT   |
|------------------|---|--|--|---|--|
| <i>Không xét</i> | <ul style="list-style-type: none"> <li>Cấu trúc đơn giản, hạn chế/ chính xác;</li> <li>Có thể thể hiện ý;</li> <li>Lỗi: cơ bản, hệ thống/ nhiều.</li> </ul> | <ul style="list-style-type: none"> <li>Từ, ngữ: thông dụng về chủ đề (rất) quen thuộc;</li> <li>Có thể lặp từ, ngữ;</li> <li>Đúng chính tả / dạng thức.</li> </ul>                         | <ul style="list-style-type: none"> <li>Độ chính xác: thấp/ tương đối rõ ràng/ rõ ràng / dễ hiểu;</li> <li>Người nghe: thấy khó hiểu / yêu cầu nhắc lại;</li> <li>Lỗi: âm đơn lẻ, khá nhiều.</li> </ul> | <ul style="list-style-type: none"> <li>Từ, ngữ: (rất) ngắn/ câu đã học thuộc;</li> <li>Rời rạc/ ngập ngừng/ chậm chạp;</li> <li>Có thể nói vấp, nhắc lại;</li> <li>Ý đơn giản, hầu như không diễn đạt được ý phức tạp.</li> </ul> | <ul style="list-style-type: none"> <li>Liên kết: giữa các từ, ngữ;</li> <li>Trả lời câu hỏi: không phù hợp/ tương đối phù hợp</li> <li>Phương tiện liên kết: kết từ thông dụng - “và”, “nhưng”, “vì”; có thể lặp</li> <li>Phát triển ý: hầu như không/ hạn chế, dùng từ gọi ý trong đề, liệt kê</li> </ul> |
| 3                | <ul style="list-style-type: none"> <li>Câu đơn: chính xác / ít lỗi;</li> <li>Câu phức: có thể có nhiều lỗi.</li> <li>Có thể thể hiện ý: rõ ràng.</li> </ul> | <ul style="list-style-type: none"> <li>Từ, ngữ: về chủ đề quen thuộc/ không quen thuộc;</li> <li>Đôi lúc lặp từ, ngữ;</li> <li>Lỗi: một số / nhiều; lỗi kết hợp từ, diễn đạt ý.</li> </ul> | <ul style="list-style-type: none"> <li>Từ, âm đơn lẻ: rõ ràng, dễ hiểu;</li> <li>Lỗi: âm đơn lẻ, trọng âm; một số/ nhiều</li> <li>Có nhấn trọng âm từ.</li> </ul>                                      | <ul style="list-style-type: none"> <li>Ý đơn giản: trôi chảy;</li> <li>Ý phức tạp: chậm, ngập ngừng, do phải tìm cấu trúc và từ vựng.</li> <li>Có thể phát triển câu trả lời bằng các cụm từ, câu đơn;</li> </ul>                 | <ul style="list-style-type: none"> <li>Liên kết: có liên kết giữa các ý;</li> <li>Trả lời câu hỏi: phù hợp,</li> <li>Phương tiện liên kết: một số/ đa dạng phương tiện đơn giản / phức tạp; có thể thể hiện quan hệ liên kết chưa chính xác; đôi khi lặp.</li> </ul>                                       |

|   |   |  |   |   |  |
|---|---|--|---|---|--|
|   |   |  |   | <ul style="list-style-type: none"> <li>▪ Lỗi: có thể mắc khi nói các đoạn dài; biết sửa lỗi.</li> </ul>   | <ul style="list-style-type: none"> <li>▪ Phát triển ý: liệt kê / tương đối dễ dàng;</li> <li>▪ Mở rộng ý: chưa rõ / lặp / với (những) chi tiết, ví dụ minh họa phù hợp;</li> </ul>   |
| 4 | <ul style="list-style-type: none"> <li>▪ Các dạng câu đơn: chính xác, linh hoạt;</li> <li>▪ Câu phức: một số/ khá nhiều;</li> <li>▪ Lỗi: một số, không hệ thống, có thể tự sửa, không gây hiểu nhầm.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Từ, ngữ: về hầu hết các chủ đề quen thuộc/ không quen thuộc;</li> <li>▪ Lặp từ, ngữ: ít khi, với các chủ đề không quen thuộc;</li> <li>▪ Kết hợp từ, diễn đạt: tương đối chính xác;</li> <li>▪ Chính tả / dạng thức từ: (tương đối) chính xác (cao);</li> <li>▪ Có từ, ngữ ít thông dụng; cụm từ cố định, thành ngữ.</li> <li>▪ Lỗi: một số; lỗi kết hợp từ, dạng thức từ, diễn đạt ý.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Từ, âm đơn lẻ: (hầu như) rõ ràng, chính xác, tự nhiên, dễ hiểu;</li> <li>▪ Trọng âm từ: tương đối chính xác;</li> <li>▪ Có nhấn trọng âm câu;</li> <li>▪ Có thể hiện ngữ điệu câu.</li> </ul>  | <ul style="list-style-type: none"> <li>▪ Ý đơn giản và phức tạp: giao tiếp dễ dàng, tương đối/ khá lưu loát;</li> <li>▪ Có thể nói các đoạn dài: có thể lặp từ;</li> <li>▪ Lỗi: có thể mắc một số lỗi nhỏ/ không hệ thống, có thể sửa lỗi;</li> <li>▪ Ngập ngừng: thỉnh thoảng, không quá lâu.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Phương tiện liên kết: nhiều/ đa dạng; thể hiện (khá) rõ ràng mối quan hệ giữa các ý.</li> <li>▪ Phát triển ý: dễ dàng;</li> <li>▪ Mở rộng ý: với những chi tiết, ví dụ minh họa phù hợp.</li> </ul>   |
| 5 | <ul style="list-style-type: none"> <li>▪ Nhiều cấu trúc câu: đa dạng, linh hoạt, chính xác;</li> <li>▪ Lỗi: khó phát hiện/ hầu như không.</li> </ul>  | <ul style="list-style-type: none"> <li>▪ Từ, ngữ: phong phú; có từ, ngữ thuộc tần số sử dụng thấp; thành ngữ;</li> <li>▪ Lặp từ, ngữ: hiếm khi; có thể còn còn ngập ngừng khi tìm từ, ngữ thay thế;</li> <li>▪ Lỡ lời: đôi khi/ hầu như không;</li> <li>▪ Lỗi: nhỏ; một vài / không.</li> </ul>  | <ul style="list-style-type: none"> <li>▪ Từ, âm đơn lẻ: rõ ràng, chính xác, tự nhiên, dễ hiểu;</li> <li>▪ Trọng âm từ, câu: chính xác; thay đổi để thể hiện các sắc thái ý nghĩa khác nhau;</li> <li>▪ Ngữ điệu câu: phù hợp; thay đổi để thể hiện các sắc thái ý nghĩa khác nhau.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Đoạn dài: dễ dàng, lưu loát, tự nhiên;</li> <li>▪ Lặp từ, sửa lỗi: đôi khi, hiếm khi;</li> <li>▪ Ngập ngừng: hầu như không; chỉ khi diễn đạt một số chủ đề/ khái niệm khó.</li> </ul>  | <ul style="list-style-type: none"> <li>▪ Phát triển, mở rộng ý: dễ dàng, (khá) tường tận, với những chi tiết, ví dụ minh họa phù hợp; kết luận phù hợp;</li> <li>▪ Tổ chức ý: rõ ràng, mạch lạc, cấu trúc chặt chẽ;</li> <li>▪ Phương tiện liên kết: (khá) thành thạo, đa dạng các phương tiện.</li> </ul> |

Table 3: Scoring rubrics for Writing test (Tiêu chí chấm điểm bài thi Viết)

| BAND                      |        | Không xét   | 3  | 4   | 5   |
|---------------------------|--------|---|--|---|---|
| MỨC ĐỘ HOÀN THÀNH BÀI THI | TASK 1 | <ul style="list-style-type: none"> <li>▪ Không đáp ứng yêu cầu nào của đề bài;</li> <li>▪ Bài viết hoàn toàn không phù hợp, không thể hiểu được;</li> <li>▪ Trình bày được rất ít ý, phần lớn những ý này không phù hợp, lặp lại.</li> </ul>  | <ul style="list-style-type: none"> <li>▪ Đáp ứng một nửa hay một phần yêu cầu của đề bài;</li> <li>▪ Thể hiện mục đích bài viết không rõ ràng (ở một số chỗ);</li> <li>▪ Có thể nhầm lẫn ý chính với thông tin chi tiết;</li> <li>▪ Một số phần có thể không rõ ràng, không phù hợp hoặc lặp lại;</li> <li>▪ Các điểm chính có thể không được trình bày đầy đủ; có xu hướng tập trung vào những chi tiết;</li> <li>▪ Có thể mắc một số lỗi về tính thống nhất, phù hợp trong giọng văn.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Đáp ứng (gần như) đầy đủ yêu cầu của đề bài;</li> <li>▪ Thể hiện mục đích viết (nhìn chung) rõ ràng;</li> <li>▪ Dạng thức bài viết có thể không phù hợp ở một số chỗ;</li> <li>▪ Các điểm chính: được trình bày (gần như) đầy đủ;</li> <li>▪ Phát triển ý: hầu hết thông tin chi tiết phù hợp; 1-2 điểm chính có thể cần phát triển hơn nữa;</li> <li>▪ Có thể mắc 1-2, một số lỗi (nhỏ) về tính thống nhất, phù hợp trong giọng văn.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Đáp ứng các yêu cầu của đề bài đầy đủ, hiệu quả;</li> <li>▪ Phát triển điểm chính đầy đủ với (tất cả) thông tin chi tiết phù hợp;</li> <li>▪ Thể hiện mục đích viết rõ ràng, hiệu quả; giọng văn thống nhất, phù hợp.</li> </ul> |
|                           | TASK 2 | <ul style="list-style-type: none"> <li>▪ Bỏ thi / Không viết từ nào/ Viết lại một bài theo trí nhớ/thuộc lòng;</li> <li>▪ Bài hoàn toàn không phù hợp/ không thể hiểu được;</li> <li>▪ Hầu như không đáp ứng yêu cầu nào của đề bài;</li> <li>▪ Quan điểm không được thể hiện;</li> <li>▪ Không trả lời đầy đủ bất cứ yêu cầu nào của đề bài;</li> <li>▪ Trình bày 1-2 ý chính nhưng (phần lớn) không được phát triển.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Đáp ứng một phần yêu cầu của đề bài;</li> <li>▪ Những ý chính: trình bày một số, có thể lặp, phát triển không đầy đủ với những chi tiết phù hợp/ không phù hợp;</li> <li>▪ Quan điểm thể hiện không hẳn rõ ràng trên toàn bài; có thể không rút ra kết luận.</li> </ul>   | <ul style="list-style-type: none"> <li>▪ Đáp ứng (gần như) đầy đủ các yêu cầu của đề bài;</li> <li>▪ Những ý chính: phát triển gần như đầy đủ, với hầu hết thông tin chi tiết phù hợp;</li> <li>▪ Quan điểm thể hiện nhìn chung rõ ràng trên toàn bài, có thể lặp/ không rõ ở một số phần/ phần kết luận.</li> </ul>  | <ul style="list-style-type: none"> <li>▪ Đáp ứng đầy đủ / hiệu quả các yêu cầu của đề bài;</li> <li>▪ Những ý chính: phát triển đầy đủ, phù hợp, được mở rộng/ chứng minh đầy đủ;</li> <li>▪ Quan điểm thể hiện rõ ràng trên toàn bài.</li> </ul>                         |



| TỔ CHỨC BÀI VIẾT | <ul style="list-style-type: none"> <li>▪ Phương tiện liên kết những từ nối cơ bản thường gặp;</li> <li>▪ Tổ chức thông tin, lập luận: không có, hạn chế.</li> </ul>   | <ul style="list-style-type: none"> <li>▪ Chia đoạn không hợp lý, hoặc không viết theo đoạn;</li> <li>▪ Phương tiện liên kết: sử dụng đúng một số từ nối, phương tiện cơ bản, thường gặp, ở trong câu, liên câu; có thể lặp, thiếu chính xác;</li> <li>▪ Tổ chức thông tin, lập luận: mạch lạc.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Chia đoạn (tương đối) logic;</li> <li>▪ Phương tiện liên kết: trong câu, liên câu; nhiều; phù hợp; có thể quá nhiều/ ít/ đôi khi còn sai;</li> <li>▪ Tổ chức thông tin, lập luận: hợp lý, mạch lạc.</li> </ul>           | <ul style="list-style-type: none"> <li>▪ Chia đoạn đủ và hợp lý;</li> <li>▪ Phương tiện liên kết: phong phú, linh hoạt, hiệu quả;</li> <li>▪ Tổ chức thông tin, lập luận: lô-gíc, mạch lạc, hiệu quả.</li> </ul>   |
|------------------|---|---|---|--|
| TỪ VỰNG          | <ul style="list-style-type: none"> <li>▪ Sử dụng phạm vi từ (rất) hạn chế.</li> <li>▪ Sử dụng từ, ngữ rời rạc;</li> <li>▪ Mắc (rất) nhiều lỗi, lỗi có thể làm thay đổi nghĩa;</li> </ul>  | <ul style="list-style-type: none"> <li>▪ Sử dụng phạm vi từ hẹp theo chủ đề bài viết; có xu hướng lặp một số từ;</li> <li>▪ Sử dụng từ, ngữ cơ bản;</li> <li>▪ Mắc (khá nhiều) lỗi, thỉnh thoảng gây khó hiểu.</li> </ul>   | <ul style="list-style-type: none"> <li>▪ Sử dụng được dải từ rộng</li> <li>▪ Sử dụng một số từ ít thông dụng, ngữ cố định, có thể mắc lỗi;</li> <li>▪ Dùng từ tương đối phù hợp phong cách;</li> <li>▪ (Rất) Ít lỗi hệ thống, lỗi có thể gây khó hiểu.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Sử dụng dải từ rất rộng, bao gồm những từ ít thông dụng, ngữ cố định; chính xác, linh hoạt;</li> <li>▪ Dùng từ (nhìn chung) phù hợp phong cách;</li> <li>▪ (Hầu như) không mắc lỗi, hoặc 1-2 lỗi không hệ thống; không gây khó hiểu.</li> </ul> |
| NGŨ PHÁP         | <ul style="list-style-type: none"> <li>▪ Viết một vài từ, ngữ đã học thuộc trước;</li> <li>▪ Không viết được thành câu;</li> <li>▪ Cấu trúc đơn giản: một vài, sử dụng chính xác;</li> <li>▪ (Nhiều) lỗi cơ bản, làm thay đổi nghĩa;</li> <li>▪ Lỗi xuất hiện rất nhiều và làm thay đổi nghĩa.</li> </ul> | <ul style="list-style-type: none"> <li>▪ Cấu trúc đơn giản: sử dụng chính xác;</li> <li>▪ Cấu trúc phức tạp: có thể sử dụng sai;</li> <li>▪ Có thể (tương đối nhiều) lỗi, nhưng không, hoặc thỉnh thoảng gây khó khăn cho người đọc;</li> </ul>   | <ul style="list-style-type: none"> <li>▪ Cấu trúc: đơn giản và phức tạp;</li> <li>▪ Đa số các câu không có lỗi;</li> <li>▪ Có thể mắc lỗi không hệ thống, hầu như không gây khó hiểu, hiểu nhầm.</li> </ul>   | <ul style="list-style-type: none"> <li>▪ Cấu trúc: đa dạng - đơn giản và phức tạp; chính xác, linh hoạt;</li> <li>▪ Gần như không mắc lỗi, hoặc 1-2 lỗi không hệ thống, không gây khó hiểu.</li> </ul>   |

Table 4: The original scoring rubrics vs. The revised

| Features        | SPEAKING scoring rubric |            | WRITING scoring rubric     |            |
|-----------------|-------------------------|------------|----------------------------|------------|
|                 | ORIGINAL                | REVISED    | ORIGINAL                   | REVISED    |
| No. of column   | 12                      | 6          | 10 (five each)             | 5          |
| No. of row      | 6                       | 5          | 24 (12 each)               | 9          |
| No. of word     | <b>2,230</b>            | <b>856</b> | <b>2,682</b> (1259 + 1423) | <b>931</b> |
| No. of page     | <b>5</b>                | <b>2</b>   | <b>10</b> (5 each)         | <b>2</b>   |
| No. of levels   | <b>11</b>               | <b>4</b>   | <b>11</b>                  | <b>4</b>   |
| No. of features | 5                       | 5          | 4                          | 4          |

Luoma [6, p.80] maintains the number of levels in a scale is one of the ‘central questions in scale development’, adding that ‘the lower the number of levels, the more consistent the decisions.’ The basic rules to have been followed when I revised the descriptors are “brief, clear, definite, and comprehensible independently” [4, p. 205-207]. Brevity makes these scales user-friendly, for assessors, who may use the scales as a reference during the rating process, and for assessor-trainers, who need to facilitate the trainees in training workshops. Clarity refers to simple wordings and sentence structures to help readers understand the descriptors more quickly and easily.

In order to assure scoring quality, every effort should be made to maintain high reliability in assessing these two productive skills, such as:

- Conducting seminars on assessment, VSTEP, CEFR and/or Vietnamese 6-level framework, when the trained, experienced assessors returning from workshops can train the others in the staff;

- Regarding the scoring rubrics, there should be extensive training so that would-be raters come to an understanding of the criteria through repeated discussion and exposure to

multiple sample scripts/ oral performances that exemplify points on the scale. It is believed that “only when all raters are in close agreement on the scores for these samples that reliable scoring can take place.” (White, 1984, cited in [5, p. 128])

- As far as the speaking test is concerned, each candidate’s performance must be recorded for the second independent rating.

- As for the writing test, there can be a number of other measures. First, each script must be scored independently by two raters. The first rating can be done on a rating sheet, which is hidden from the second rater; they are allowed neither to write any comments nor to underline errors when scoring scripts. A third rater is invited to adjudicate in cases of discrepancy. Second, rating must be done in a “controlled reading” - “readers meet together to grade scripts at the same place and time” in order to eliminate unnecessary sources of error variance and to form a positive social environment helping to maintain the assessing standards (White, 1984, cited in [5, p. 128]). An additional measure is randomly checking the results by senior assessors to ensure individual raters have followed the agreed-upon standards for grading of the ongoing assessment. This step is also aimed to keep records on the level



of expertise of the raters so that decisions can be made on whether each can be kept on, retrained, or dropped.

#### 4. Conclusion

In this article, to my best knowledge, for the first time, the scoring rubrics of the VSTEP 3-5, is critically analyzed and described from the perspective of Vietnamese grammar and a revised rubric is proposed, which is far more concise, linguistically grounded, and undoubtedly useful to users.

Weigle (2020) states:

*the scoring rubric is critical, as it represents as explicitly as possible the definition of the skill(s) that the test is intended to measure...It is not enough for a rubric to be clear and explicit: it must also be useable and interpretable, certainly by raters, and preferably by any and all stakeholders in the testing process, particularly test takers and decision makers. (p. 122)*

VSTEP 3-5 has had a positive reception among students, lecturers, and other community members. I hope that the revised versions are of some practical significance to the endeavor of the stakeholders in order to bring excellence to EFL education and assessment.

#### References

- [1] Bộ Giáo dục và Đào tạo, Quyết định số 729/QĐ-BGDĐT “Ban hành định dạng đề thi đánh giá năng lực sử dụng tiếng Anh từ bậc 3 đến bậc 5 theo khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam, Hà Nội, ngày 11 tháng 3 năm 2015”.
  - [2] Quynh, N. T. N. (2020), Vietnamese Standardized Test of English Proficiency - A Panorama. In Lily I-Wen Su, et al. (eds). (2020), *English Language Proficiency Testing in Asia*. Routledge, pp. 71-1000
  - [3] Bộ Giáo dục và Đào tạo, Tài liệu “Hướng dẫn áp dụng Định dạng đề thi đánh giá năng lực sử dụng tiếng Anh từ bậc 3 đến bậc 5 theo Khung năng lực ngoại ngữ 6 bậc dùng cho Việt Nam trong việc xây dựng đề thi và chấm thi (Phê duyệt kèm theo Quyết định số: 730/QĐ-BGDĐT ngày 11 tháng 3 năm 2015 của Bộ trưởng Bộ Giáo dục và Đào tạo).
  - [4] Council of Europe. 2001. The Common European Framework of Reference for Languages (CEFR) - euroexam international ([www.euroexam.org](http://www.euroexam.org))
  - [5] Weigle, S. C. 2002. *Assessing Writing*. New York: Cambridge University Press.
  - [6] Luoma, S. 2002. *Assessing Speaking*. New York: Cambridge University Press.
- KPG exams. The A/ B/ C level. Speaking Test – Assessing Oral Production. Oral Examiners information Pack, September 2014. [https://rcel2.enl.uoa.gr/kpg/files/C\\_level\\_oral\\_exam\\_iner\\_info\\_pack.pdf](https://rcel2.enl.uoa.gr/kpg/files/C_level_oral_exam_iner_info_pack.pdf)
- Pölzleitner and Bauer. 2020. Assessment Scale for Written Work / Assessment Scale for Written Work in Lower School (<https://www.polzleitner.com/epep/Assessment/NewAssessmentScales/all-three-files.pdf>).