Weighted six-trait writing scale by native and non-native writing instructors: rank of importance

Ceylan Yangın Ersanlı[1]

Ondokuz Mayıs University, Samsun-Turkey

Abstract

This study presents the findings of a descriptive study addressing the question of whether writing evaluations of non-native English speaking (NNS) instructors match with those of native English speaking (NS) instructors on the basis of a six-trait writing rubric. Given the fact that several Schools of Languages at tertiary level in Turkey hire native English speaking teachers to optimize the learning conditions for the learners and that they carry out the courses with their non-native colleagues complementarily, it is worth comparing native and non-native English instructors' grading criteria of writing. The present study explores evaluations of 30 writing instructors (15 Native and 15 Non-native instructors) and the magnitude of weights assigned by them in terms of a six-trait writing rubric (ideas & content, organization, voice & tone, word choice, sentence fluency and conventions). The writing rubric is developed based on the related literature. Both groups of instructors were asked to rank the sections of the rubric according to their relevant importance. The results showed that both groups value ideas and content as the most important trait which is followed by organization. The NS instructors assign more weight to sentence fluency while the NNS instructors value word choice more. Both groups assign more weight to conventions of writing than voice and tone.

Keywords: Magnitude of weights, native and non-native instructors, writing evaluation

---

[1] Ondokuz Mayıs University, Faculty of Education, Department of English Language Teaching.
Email: ceylanyangin@gmail.com

Introduction

In the area of language teaching and learning, teachers' evaluation criteria regarding the students' performances will inevitably affect their teaching styles, the methods and the activities they prefer to use. Therefore, most scholars such as Chastain (1980) would agree that evaluation is central not only to assess students' performances but also to give them accurate and appropriate feedback and to determine the most compatible approach to instruction for the students. For this reason, an objective ranking basis has to be established among evaluators. However, there may be some significant differences in teachers' evaluations in such areas as linguistic accurateness, comprehensibility, communicative appropriateness, and so on. Some teachers may tend to be more severe in their evaluations on particular areas whereas the others may ignore or tolerate them. The underlying differences among the raters' evaluations may arouse from many reasons; the raters may come from different linguistic or professional backgrounds, they may have different evaluation cultures, or they may have different instructional goals. For instance, if the teacher determines his goal as communicative success rather than linguistic accurateness, he may ignore or tolerate the linguistic errors in the students' performances, or the vice versa.

In recent years the promotion of native speaker-norms in language classrooms has two facets; enrolling native speaker teachers in the language teaching programs, and the increasing demand for language proficiency tests such as TOEFL (Test of English as a Foreign Language) and IELTS (International English Language Testing System). These tests assess students' performances using native speakers as a benchmark (Lowenberg, 2002). However, English is becoming a lingua franca, a world language. It is used in international contexts as a contact language by people from diverse cultures and languages. This requires a reconsideration about the role of NS as the only acceptable standard as language assessors. In Turkey and in most non-English speaking countries English is taught as a foreign language not as a second language. Therefore, students have limited exposure to target language input. Therefore, enrolling NS teachers into language programs has risen in prominence. However, the difference between NS teachers and NNS teachers' cultural and linguistic backgrounds will inevitably allow variations in their rating criteria. Another vital reason behind the differences between NS and NNS teachers' evaluations is their professional backgrounds. Most of the NS teachers working abroad may be from different disciplines such as anthropology, politics, art history and so on. They may not be the graduates of faculty of education. These variations between the teachers may cause different rating criteria. This is evident especially in the writing assessment since as Robinson (2000, p. 667) states in writing "…no such broadly accepted instruments or scoring mechanisms exist". Other scholars such as Cohen (1994), Scott and Rodgers (1993) concur with the idea that in writing assessment less agreement exists as what constitutes a reliable rating for writing.

In order to build a consensus on the NS teachers and NNS teachers' evaluations of the students' writing performances, first the goal of the instruction should be determined: is it linguistic accurateness or communicative appropriateness that should matter most? By this way a basis for evaluation can be established and the results can be applied to language instruction, material selection, and the activities to be used.

Related Research

Recently, many researches regarding the rater variability have focused on the NS and NNS teachers' evaluation norms. Studies have been conducted to explore the effect of teachers' cultural, linguistic and professional backgrounds on the assessment of students' written or oral target language performances by asking both NS and NNS teachers evaluate learner language samples.

The general idea is that NS teachers and non-teachers give more importance to communicative aspects of learner language which means if a learner language is comprehensible it is acceptable. On the other hand, it is generally believed that NNS teachers put more emphasis on the conventions 'the linguistic aspects' of the language produced neglecting the communicative value of it. They are seen more severe in their evaluations in terms of the accurateness of the language ignoring the appropriateness dimension. Hyland and Anan (2006, 511) concurs with this idea by stating that "[g]enerally, findings show non-native speakers to be more severe, to be more obviously oriented accuracy…". However, while the findings of some studies support this belief (Fayer and Krasinski, 1987; Galloway, 1980) in other studies dissimilarities emerged. For example, Shi (2001) investigated the writing performances of students. The study focused on the writings of Chinese learners of English. It revealed that NNS teachers tended to identify more negative features in learners' writing whereas the NS made more positive comments. However, in terms of scoring the learners' papers, the ratings were vice versa; NS raters assigned lower marks than the NNS teachers (cited in Johnson and Lim, 2009).

Kim (2009) investigated if there existed any difference between NS and NNS teachers' assessment of students' English speaking performances. The results suggested that there were dissimilarities in the evaluation criteria. NS teachers were seen as more critical and to the point in terms of pronunciation and accuracy in students' oral performances. However, in a similar study Zhang and Elder (2011) drew on different constructs. They tried to address the question of whether NNS and NS teachers' evaluations of the oral performances of students were similar or not. The results indicated that although there were no differences in the sores assigned by both of the groups, NNS teachers appeared to be more accuracy-focused and less meaning-focused than NS teachers.

Robinson's study (2000) focused on the writing performances of students and sought to discover whether NNS teachers and non-educator native informants would rate students' writing in a similar fashion. Results demonstrate that when errors of meaning and form occur, both groups are consistent and score the sample as below proficient. When some errors of form are present yet meaning is clear, teachers tend to be more forgiving and rate the sample as proficient. The non-educator native informants found the surface errors to be more distracting, which resulted in a rating of below proficient.

Furneaux, Paran and Fairfax (2007) investigated to what extent NNS teachers from different linguistic backgrounds differ in evaluation of students' writings. They tried to find answers to questions such as do NNS teachers from different linguistic backgrounds take on different roles?, which aspects of student writing do teachers focus on? Results suggested that

teachers have similar roles while evaluating the students' essays and they responded to the essays as readers of communication.

Johnson and Lim (2009) focused on the question whether rater's language background had an influence on writing assessment. To this end, both NS teachers of English and NNS teachers of English from different countries participated into the study. Results put forward that there was no difference in the ratings of these groups.

In Hyland and Anan's study (2006) three different groups were investigated in terms of their beliefs and practices: NS teachers, NNS teachers and educated NS non-teachers. Each group identified and corrected errors in a student's written work. The NNS teachers tended to be critical in evaluating the errors focusing more on accurateness than communicativeness, whereas the NS teachers were more cautious about formal and academically appropriate language use.

Based on these studies, it can be suggested that there is a connection between the evaluations of learner language samples and evaluators' linguistic, and professional background. However, the studies on how NS and NNS teachers evaluate learners' target language performance are relatively few and no consensus has been reached. Thus, further research is needed especially in writing assessment for two reasons. First, most of the studies on the evaluators' linguistic or professional background have been on the oral production of the learners. As Johnson and Lim (2009) stated, it is worth investigating whether a rater's background effects assessment of writing. Second, teaching and assessing writing in a foreign language has changed very little over the years. It has still been developing focusing on writing as a communicative skill. The findings of such studies will shed a light on the discussion of the degree of emphasis given to linguistic aspects of a piece of writing and its content, communicative value.

It is believed worthwhile to investigate whether evaluations of NS and NNS instructors differ with respect to assessing written work. Although a well-defined writing rubric is used, both groups of instructors may differ in their emphasis in the sub-sections. To this end, this study addresses the following research questions:

1) How do NS instructors rank sections of a six-trait writing rubric in terms of importance?
2) How do NNS instructors rank sections of a six-trait writing rubric in terms of importance?
3) Do NS instructors and NNS instructors differ in their in their emphasis on the sections of writing rubric grading criteria?

Methodology

Design

The current study is a descriptive one, thus adopting a quantitative research type including the weighting of a writing rubric by both NS and NNS instructors.

Participants

A total of 30 instructors participated in the study (15 native speaker instructors and 15 non-native instructors), all of whom had instructed academic writing at the tertiary level in preparatory classes in the state and private universities in Turkey. All of the participants asserted that they had taught writing as a separate course or as part of the whole program.

Table 1

Native and Non-Native Writing Instructors Participating in the Study

| | Gender | | Teaching Experience | | | | English Major | |
	Male	Female	Up to 4 years	4-8 years	9-12 years	13 years +	Yes	No
Native	5	10	11	2	2	-	5	10
Non-Native	6	9	4	5	4	2	15	-

Most of the NS participants were novice teachers with only a few experienced teachers. Their teaching experience ranged from 2 months to 12 years. Except from two experienced NS teachers who had teaching experience of 11 and 12 years, most of the NS instructors had had experience in Turkish education context ranging between 2 months to 4 years (six instructors with a teaching experience of two months, three with 2 years of experience and two with four years of experience). The rest had relatively more teaching experience in the given context (five and six years of teaching experience). As to their educational background, two thirds of them did not major in English language. To be precise, three were the graduates of Anthropology, two had a degree in Political Sciences, one in Business Administration, one in Engineering, one in Art History, one in Biology and one in Social Sciences whereas five of them majored in English language and literature. None of the NS instructors had an M.A. degree.

Most of the NNS instructors' teaching experience ranged from 4 to 15 years. Only four NNS instructors had less experience than four years. Yet, of these the one with the least teaching experience had been working at the tertiary level language programs for two years. In terms of experience in teaching, the NNS group might be regarded as a lot more experienced than their NS counterparts. All of the NNS instructors had a degree in the English Language Teaching Departments in Turkey. Only four of them had an MA degree.

Research

The study aims to explore decisions of two groups of instructors to weigh criteria. Specifically, the study attempts to determine the magnitude of weights assigned to subsections of a six-trait writing rubric by NS instructors and NNS instructors so as to explore whether the two groups display different evaluation criteria for writing. First, a writing rubric was developed based on the related literature. The content validity of the rubric was ensured by four professionals in the areas of Language Teaching and Language Assessment. The respondents were asked to distribute 100 points among the sections in the given writing rubric

to see how much each section weighs in their opinion. By doing so, the participants were asked to rank the subsections of the writing rubric in order of importance. They were also informed that they may assign no points to the sections that they believe unnecessary or dispensable, or assign the same weight to different sections if they believe they are of equal importance. They were also asked to include any other section that they would like to add.

The magnitude of weights is evaluated as follows:

The participants are asked to rank the subsections of the writing rubric according to their perceived importance from 1 as the most important and 6 as the least important. They may also assign the same rank to two or more sections if they believe they are equally important. Then, for each section, the number of participants assigning different weights is noted down. The magnitude of weights is calculated by multiplying the number of participants by the weight for each section (the first rank receives the highest point possible-6, the second rank is multiplied by 5, the third by 4… and the last rank is multiplied by 1). The sum accounts for the total points each section is granted by the participants. The sums for each section are then summed up to get the total weight for the whole rubric. The ratio of the sums of each section to the total weight yield the percentage, so that the subsections of the writing rubric can be compared in terms of importance attached to them by the participants.

An example of the calculation is given below:

To calculate relevant importance attained to the sub-section "organization" by NS instructors (see table below), the 3 persons assigning the first rank were multiplied by 6 points, 11 participants assigning the second rank by 5 points and 1 person assigning the fifth rank with 2 points: (3x6)+(11x5)+(1x2)=75. After the sums were added up for each section, the total points were attained (373 for NS instructors). The ratio of each section to the total score yielded the percentages for comparable results.

Findings and Discussion

The following table displays the magnitude of weights assigned by NS instructors for each of the sections in the six-trait writing rubric.

Table 2

Magnitude of Weights by NS Instructors

Rank / points	1 (6pts)	2 (5pts)	3 (4pts)	4 (3pts)	5 (2pts)	6 (1pt)	sum	%
Ideas and Content	14		1				88	23,6
Organisation	3	11			1		75	20,1
Voice and Tone		1	1	6	6	1	41	11
Word Choice		3	3	6	5		55	14,7
Sentence Fluency		4	6	2	1	2	57	15,2
Conventions	3	4	2	1	3	2	57	15,2
TOTAL							**373**	

As can be seen, *Ideas and Content* account for the most important aspect in the evaluation of an essay for NS instructors (23,6%). It appears that developing clear and focused ideas to enrich the theme is of greatest importance. The theme or argument along with the supporting details constitutes the core of essays for the NS instructors. Following that, *Organisation* trait is regarded as the second most important aspect of writing (20,1%). This is related with how a piece of writing is structured in a well-planned introduction, body and conclusion. The NS instructors prioritize the order of ideas as well. *Sentence Fluency* and *Conventions* receive the same weighting by the NS instructors (15,2%). The length and variety in sentences and accuracy and mechanical correctness in spelling and punctuation have the third place in weighting. *Word Choice*, that is the use of sophisticated vocabulary appropriately, is regarded as the next important trait in the rubric (14,7%). The NS instructors place *Voice and Tone* in the last rank of order (11%). The emotional attitude expressed or the formality or casuality of the writing receives relative less importance compared to the other five traits.

Table 3

Magnitude of Weights by NNS Instructors

Rank / points	1 (6pts)	2 (5pts)	3 (4pts)	4 (3pts)	5 (2pts)	6 (1pt)	sum	%
Ideas and Content	11	3				1	82	22
Organisation	10		1	1	3		73	19,6
Voice and Tone	1		2	7	1	4	39	10,5
Word Choice	6	1	5	3			70	18,8
Sentence Fluency	2		6	3	2	2	51	13,7
Conventions	4	3	1	3	1	3	57	15,3
TOTAL							**372**	

As to NNS instructors, Ideas and Content hold the first place in weighting (22%). The instructors prioritize clear, complete and well-developed ideas with adequate support. Organisation, that is the thread of main idea and patterns of supporting ideas and transitions, weight the second most important (19,6%). The NNS instructors regard Word Choice in the third place in magnitude of weights (18,8%). Challenging vocabulary and using the right words in the right way is considered as an important aspect of essays. Word choice is followed by *Conventions.* Grammatical correctness and aspects such as punctuation and capitalisation weigh 15,8% in weighting by NNS instructors. *Sentence Fluency* receives the fifth rank (13,7%). Craftmanship in sentences is regarded as relatively less important that word choice by the NNS instructors in writing. Finally, *Voice and Tone* is assigned the sixth plave in weighting (10,5%).

Table 4

Comparing NS and NNS Instructors' Weighting of Six-Trait Writing Rubric

NS Instructors		%	NNS Instructors		%
Ideas and Content	1	23,6	Ideas and Content	1	22
Organisation	2	20,1	Organisation	2	19,6
Sentence Fluency	3	15,2	Word Choice	3	18,8
Conventions	3	15,2	Conventions	4	15,3
Word Choice	4	14,7	Sentence Fluency	5	13,7
Voice and Tone	5	11	Voice and Tone	6	10,5

When subjective weighting of the writing criteria by both groups is compared, it is apparent that the first two ranks are the same. Both NS and NNS instructors regard ideas and supporting details and the way they are organized as the most important aspects of a piece of writing. However, NS instructors prioritize sentence fluency while NNS instructors regard word choice as more important. Both groups view conventions, accuracy in grammar and mechanics, as more important than voice and tone.

Conclusion

This study attempts to explore magnitude of weights assigned by NS and NNS instructors to the sections of a six-trait writing rubric. It involves the subjective views of both groups about the relative ordering of the traits in terms of importance. It is worth comparing the views of NS instructors working in Turkish universities with Turkish co-instructors to see whether the weights they assign to the same writing rubric differ.

Both NS and NNS instructors value clear, well-developed and supported ideas as the heart of writing. The way the ideas are organized into an essay that fits the type and purpose of writing with a consciously planned and engaging introduction, body and conclusion is significant as well. As to the third rank in weighting, the NS instructors believe effective variation in sentence patterns is more important whereas NNS instructors value sophisticated word choice more. For NS instructors conventions of writing such as grammatical accuracy and mechanical correctness of writing such as punctuation, spelling and capitalization are as of the same weight as sentence fluency. On the other hand, NNS instructors prioritize conventions over sentence fluency. For both groups, the emotional attitude expressed or sounding formal or casual, distant or intimate depending on the audience and purpose weighted as the last trait.

As the number of participants were too few for comparative statistics, the gender and educational background of participants were presented as descriptive statistics. This may be regarded as a limitation of the study. As for further research, actual student essays might be graded by NS and NNs instructors using the same writing rubric to explore how they grade the sub-sections of the rubric. Differences in views about relative weightings of traits might also cause Ns and NNS instructors to assign different grades for the same student essays.

References

Chastain, K. (1980). Native speaker reaction to instructor-identified student second-language errors. *The Modern Language Journal*, 64(2), 210-215.

Cohen, A. (1994). *Assessing language ability in the classroom*. Boston: Heinle and Heinle.

Fayer, J. M. & Krasinski, E. (1987). Native and nonnative judgements of intelligibility and irritation. Language Learning, 37, 313-326.

Furneaux, C., Paran, A. & Fairfax, B. (2007). Teacher stance as reflected in feedback on student writing: An empirical study of secondary school teachers in five countries. *International Review of Applied Linguistics in Language Teaching,* 45, 69-94.

Galloway, N. B. (1980). Perceptions of the communicative efforts oıf American students of Spanish. *Modern Language Journal*, 64, 428-433.

Hyland, K. & Anan E. (2006). Teachers' perceptions of error: The effects of first language and experience. *System*, 34, 509-519.

Johnson J. S. & Lim G. S. (2009). The influence of rater language background on writing performance assessment. *Language Testing*, 26(4), 485-505.

Kim, Y. (2009). An investigation into native and non-native teachers' judgements of oral English performance: a mixed method approach. *Language Testing*, 26(2), 187-217.

Lowenberg, P. (2002). Southeast Asian Norms: Implications for assessing proficiency in English as a global language. Paper presented at the AILA Conference, Singapore, December.

Robinson, D. W. (2000). Building consensus on the scoring of students' writing: A comparison of teacher scores versus native informants' scores. *The French Review*, 73(4), 667-688.

Scott, R. & Rogers, B. (1993). Assessing communication in writing: The development of a Spanish writing contest. *Foreign Language Annals*, 26(3), 382-392.

Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing*, 18, 303-325.

Zhang, Y. & Elder, C. (2011). Judgements of oral proficiency by non-native and native English speaking teacher raters: competing or complementary constructs?. *Language Testing*, 28(1), 31-50.