TEACHERS' PROBLEMS AND SOLUTIONS IN ASSESSING STUDENTS' WRITING IN SENIOR HIGH SCHOOL LEVEL: AUTHENTIC ASSESSMENT OR TRADITIONAL ASSESSMENT

RONAULI SIHOMBING

Citra Cemara School
Bandung, e-mail: ronauli_sihombing@yahoo.com

Abstract

English is one of the core subjects at junior and senior high school in Indonesia. Writing as the productive skill of learning English claims that the teachers have to involve in designing the writing task or assignment and designing or adapting the scoring rubric. Assessing students' writing is nota simple task. The following research studies about the teachers' problems in assessing students' writing at the senior high school level. The research was conducted for the reason to investigate a further information about teachers' problemsand explore the solutions in assessing the students' writing. The research was undertaken at a well-known public senior high school in West Java, Indonesia. This research employs a qualitative research. A semi structured interview was used to collect the data by interviewing the English teachers in the school. The data in the form of extended text were analyzed and interpreted to get the final results. Assessing writing is a complex task for teachers. It was found that there are some problems faced by the teachers in assessing students' writing. One of which is that teachers doesn't have sufficient time to correct the students' writing. Therefore, the authentic assessment is difficult to be implemented. It is recommended that teachers have to be aware of these problems because teachers' judgement on students' writing can become a crucial feedback for students to know how well they can do on writing.

Keywords: *Assessing writing, authentic assessment, teachers' problem, traditional assessment*

1. Introduction

The ability to write a second or foreign language is recognized as an important skill for educational, business and personal reasons (Weigle: 2009). In addition, other cited studies show that writing assignments can enhance students'critical thinking skills (Grauerholz 1999; Malcom 2006), help them to reflect on the origin of their beliefs (Hudd and Bronson 2007), and foster their "sociological imagination" (Edwards, 2002; Roberts, 1993). In line with these ideas, the ability to write is an important part in our community for both academic field and daily

life. The best way to test students' writing ability is to get them to write by considering at least two basic components: instructions that tell test takers what to write as well as assessment to evaluate the writing samples produced by the test takers (Hughes, 1989:75).The most appropriate way to assess writing proficiency is to have people write one or more texts (Huot, 1990).

Writing can be time consuming and difficult to teach, many teachers may not feel qualified and less of confidence in their teaching of writing and they often avoid teaching writing skills, because they do not feel comfortable with writing (Shin, 2003). Both teaching and assessing writing are difficult. The assessment of writing is no simple task as the raters (teachers) need to consider students' writing ability and what is to test: hand writing ability, correct spelling, correct grammatical sentence, paragraph construction or logical development of a main idea? (Brown, 2010).All these measurement factors will cause text quality between one student is different with other students. Besides that teacher has to be aware of the way they assess students' writing. Heterogeneity of items such as spelling, writing speed, capitalization, punctuation, and writing quality also can present a special challenge in determining a scale's dimensionality (Erford et al. ; 2001).

There are three types of rating scales in assessing writing that can be developed by teachers as judgements of their scoring criteris: primary trait scales, holistic scales, and analytic scales (Weigle, 2009). The first, primary trait scoring, in Primary trait scoring, the rating scale is defined with respect to the specific writing assignment and essays are judged according to the degree of success with which the writer has carried out the assignment. The example of primary trait scoring guide in is shown in figure 1. The second, holistic scoring, different from primary trait scoring, holistic scoring is the assigning of a single score to a script based on the overall impression of the script. Holistic scoring has been widely used in assessing writing because of its practicality. Other advantage of holistic scoring is the writers are rewarded for what they do well (White in Weigle, 2009). A well-known example of a holistic scoring rubric in ESL is the scale used for the

TOEFL Writing Test (see figure 2). The third, analytic scoring, in analytic scoring, scripts or students' writing are rated on several aspects of writing or criteria, such as content, organization, cohesion, register, vocabulary, grammar or mechanics rather than given a single score. Analytical scoring schemes provide more detailed information about a test taker's performance in different aspect of writing and are for this reason raters / teachers prefer using holistic scoring. One of the best known and most widely used analytical scales was created by Jacobs et al. (1981) (see figure 3).

These three types of rating scale can be used in assessing students' writing both in classroom assessment and large – scale assessment. In contrast, classroom teachers tend to be more concerned with other aspects of test usefulness: namely, construct validity, authenticity, instructiveness, and impact. Before designing writing assessment task or scoring procedures, we need to consider a number of key questions (Weigle, 2009). These key questions are: (1) What are we trying to test? (2) Why do we want to test writing ability? (3) Who are our test takers? (4) ho will score the tests, and what criteria or standards will be used? (5) who will use the information that our test provides? (6) What kind of information we can collect about test takers' writing ability? (7) What do we need to know about testing to make our test valid and reliable? Weigle also proposed that scoring procedures for writing assessment are critical because the score is ultimately what will be used in making decisions and inferences about writers.

Scoring criteria channels the way in which raters perceive and evaluate concrete samples of langauge performance,and finally, come to assign scores to examinees (McNamara, 1996). After designing the scoring rubric, raters also need to make writing scale descriptors. The descriptors for the various levels of the scale itself can be written. But the potential problem by using scale descriptors tends to make imprecise distinction between the levels ('exellent', 'very good', 'good', and so on). The final task in scoring procedures for writing assessment is calculating total scores. If the total score will be derived from individual raters'

score, it is assumed that two raters will read and score each script independently of each other. When two raters are in agreement, the reported score can be the sum of the average of the two raters' score. But, rating scales commonly used in assessing writing have been criticized for a number of reasons. The first criticism is that they are usually intuitively designed and therefore often do not closely enough represent the features of candidate discourse. The criteria uses impressionistic terminology brings both subjective interpretations and less precise descriptions of the nature of performance at each level (Brindley, 1998).

Figure 1. Primary Trait Scoring Guide (Lloyd-Jones, 1977)
Directions: Look carefully at the picture. These kids are having fun jumping on the overtuned boat. Imagine you are one of the children in the picture. Or if you wish, imagine that you are someone standing nearby watching the children. Tell what is going on he or she would tell it. Write as if you were telling this to a good friend, in a way that expresses strong feelings. Help your friend FEEL the experience too. Space is provided on the next three pages.

NAEP Scroing: Children on Boat

Background
Primary Trait. Imaginative Expression of Feeling through Inventive Ellaboration of a point of view.
Final Scoring Guide
ENTIRE EXERCISE
- 0 No response, sentence fragment
- 1 Scorable
- 2 Illegible or illiterate
- 3 Does not refer to the picture at all
- 4 I don't know

USE OF DIALOGUE
- 0 Does not use dialogue in the story
- 1 Direct quote from one person in the story. The one person may talk more than once.When in doubt whether two statements are made by thesame person or different of people, code 1. A direct quote of a thought also counts. Can be in hypothetical tense.
- 2 Direct quote from two or more persons in the story

POINT OF VIEW
- 0 Point of view cannot be determined, or does not control point of view.
- 1 Point of view is consistently one of the five children. Include "If I were one of the children..." and recalling participation as one of the children.
- 2 Point of view is consistently one of an observer. When an observer joins the children in the play, the point of view is still "2" because the observer makes a sixth person playing. Include papers with minimal evidence even ehen difficult to tell which point of view is being taken.

TENSE
- 0 cannot determine time, or does not control tense. (One wrong tense places the paper in this category, except drowned in the present.)
- 1 Present tense – past tense may also be present if not part of the "main line" of the story.
- 2 Past tense – If a past tense description is acceptable brought up to present, code as "past." Sometimes the present is used to create a frame for past events. Code this as past, since the actual description is, in the past.
- 3 Hypothetical time – Papers written entirely in the "If I were on the boat" or "If I were there, I would." These papers often include future references such as "when I got on the boat I will." If part is hypothetical and rest past or present and tense is controlled, code present or past. If the introduction, up totwo sentences, is only part in past or present then code hypothetical.

Figure 2. TOEFL writing Scoring Guide

6. An essay at this level:
- Effectively addresses the writing task
- Is well organized andwell developed
- Uses clearly appropriate details to support a thesis or illustrate ideas
- Displays consistent facility in use of language
- Demonstrates syntactic variety and appropriate word choice thought it may have occasional errors

5. An essay at this level:
- May addresses some parts of the task more effectively than others
- Is generally well organized and developed
- Uses details to support a thesis or illustrate an idea
- Displays facility in the use of language
- Demonstrates some syntatic variety and range of vocabulary, though it will probably have occasional errors

4. An essay at this level:
- Addresses the writing topic adequately but may slight part of the task
- Is adequately organized and developed
- Uses some details to support a thesis or illustrate an idea
- Demonstrates adequate but possibly inconsistent facility with syntax and usage
- May contain some errors that occasionally obscure meaning

3 An essay at this level may revealone or more of the following weakness:
- Inadequate organization or development
- Inappropriate or insufficient details to support or illustrate generalizations
- A noticeably inappropriate choice of words or word forms
- An accumulation of errors is sentence structure and or usage

2. An essay at this levelis seriously flawed by oneor more of the following weakness:
- Serious disorganization or undevelopment
- Little or no detail, or irrelevant spesifics
- Serious and frequent errors in sentence structure or usage
- Serious problems with focus

1 An essay at this level:
- May be incoherent
- May be undeveloped
- May contain severe and persistent writing errors

0 A paper is rated 0 if it contains no response, merely copies the topic, is off-topic, is written in foreign language, or consists of only keystroke characters.

Figure 3. Jacobs et al.'s (1981) scoring profile

STUDENT	DATE	TOPIC
SCORE	LEVEL	CRITERIA
CONTENT	30-27 26-22 21-17 16-13	EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topi GOOD TO AVERAGE: some knowledge of subject • adequate range •limited development of thesis • mostly relevant to topic, but lacks detail FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate
ORGANIZATION	20-18 17-14 13-10 9-7	EXCELLENT TO VERY GOOD: fluent expression • ideas clearly stated/supported • succinct • well-organized • logical sequencing • cohesive GOOD TO AVERAGE: somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development VERY POOR: does not communicate • no organization • OR not enough to evaluate
VOCABULARY	20-18 17-14 13-10 9-7	EXCELLENT TO VERY GOOD: sophisticated range • effective word/idiom choice and usage • word from mastery • appropiate register GOOD TO AVERAGE: adequate range • occasional errors of word/idiom form, choice, usage but not obscured FAIR TO POOR: limited range • frequent errors of word/idiom form, choice, usage • meaning confused or obscured VERY POOR: essentially translation • little knowledge of english vocabulary, idioms, word form • OR not enough to evaluate
LANGUAGE USE	25-22 21-18 17-11 10-8	EXCELLENT TO VERY GOOD: effective complex construction • few errors of agreement, tense, number, word order/function, articles, pronouns, prepositions GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense,number, word order/function, articles, pronouns, prepositions butmeaning seldom obscured FAIR TO POOR: major problems in simple/complex constructions • frequent errors of neagtion, agreement, tense, number, word order/function, articles, pronouns, prepositions and or fragments, run-ons, deletions • meaning confused or obscured VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate
MECHANICS	5 4 3 2	EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing but meaning not obscured FAIR TO POOR: occasional errors of spelling, punctuation, capitalization, paragraphing • poor handwriting • meaning confused or obscured VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate
TOTAL SCORE:		COMMENTS:

Scoring criteria channels the way in which raters perceive and evaluate concrete samples of langauge performance,and finally, come to assign scores to

examinees (McNamara, 1996). After designing the scoring rubric, raters also need to make writing scale descriptors.The descriptors for the various levels of the scale itself can be written. But the potential problem by using scale descriptors tends to make imprecise distinction between the levels ('exellent', 'very good', 'good', and so on). The final task in scoring procedures for writing assessment is calculating total scores. If the total score will be derived from individual raters' score, it is assumed that two raters will read and score each script independently of each other. When two raters are in agreement, the reported score can be the sum of the average of the two raters' score. But, rating scales commonly used in assessing writing have been criticized for a number of reasons. The first criticism is that they are usually intuitively designed and therefore often do not closely enough represent the features of candidate discourse. The criteria using impressionistic terminology brings both subjective interpretations and less precise descriptions of the nature of performance at each level (Brindley, 1998).

In writing assessment, different sources of variability like tasks, raters, and rating scales contribute to the score variance (Cooper 1984). Different rates will give different judgment for students' writing. Several studies have also examined differences between particular groups of rater in the perception and use of criteria, especially differences between experienced or expert raters and inexperienced or untrained raters (Brown *at al*., 2010). Unskilled writer are less likely to revise their spelling, punctuation, grammar, or text ideas, resulting in poorly written text (Graham & Harris, 1997; Hooper at al., 1994). There are three possible reasons why unskilled writers have shorter writing. Firstly, because students who struggle with writing terminate their writing process too soon. Secondly, Graham and Harris (1997) suggested that unskilled writers may produce shorter essays when compared to a skilled writers due to poorly developed mechanical skills. Finally, the possible reason is not related to writing skills but related to topic knowledge and interest.

There are at least three types of problematic scripts stated by (Weigle, 2009:132):

... scripts that are complete but do not address the intended task or fail to address parts of the task, scripts that have clearly been written from memory rather than in response to the prompt, and incomplete scripts – that is scripts in which the writer has demonstrated an understanding of the important features of the task but was unable to complete the task in the allotted time (for example, the conclusion may be missing.

After having a short conversation with some English teachers, they find that assessing writing is more difficult than assessing other skills. A study conducted by Cresswell (2000) found that one of the potential problems faced by the teachers regarding to assessing students' writing is when students focus on language structure rather than focus on ideas or content that make their

composition effective. The study was conducted in the area of English a first language. On the other hand, Gebril (2009), in the area of English as a second language also found the inherent problem is that if students are assigned to write a given topic without sufficient background knowledge, this variable will arise problem for the teacher to mark the students' writing because of construct - irrelevant variance. For this reason, teachers are suggested to give a source text for students such as text - based before writing that would provide them with a common platform. In line with these problems, this study aims to investigate further information about what teachers' problems are in assessing students' writing and what teachers do to enhance these kinds of problems in the area of English as a foreign language in Indonesia in Senior High School level.

Therefore, the research question of the study is what are teachers' problem in assessing students' writing? The purpose of the study is to investigate further what teachers' problems in assessing students' writingat the senior high school level. Because it is very important for students to know how far their writing ability. As McNamara stated that it involves not merely the test taker and the test,

but the test taker, the prompt or task, the written text itself, the rate(s) and the rating scale (McNamara in Weigle, 2009:108).

2. Method

Because the purposes of the study are to investigate the teachers' problems in assessing writing, the study employs qualitative research. This type of research has a greater emphasis on holistic description – that describe what all goes on in a particular activity or situation (Fraenkel e*t al.*, 2011). The research was undertaken at one of public schools (Senior High School) in Karawang, West Java in 5th and 12th December 2012. The participants of the study are two English teachers at the school. The reasons for choosing these two teachers as the participants of the research is based on the consideration that the teachers have taught English including writing for more than 10 years.

The data was collected by using interview because it is the most important data collection tecnique in qualitative research to find out what is on the interviewees' minds – what they think or how they feel about something (Fetterman & Patton in Fraenkel e*t al.*, 2011). Theaudio recorder was used to record the interview session because recording device is an indispensable part of qualitative researcher's equipment (Fraenkel e*t al.*, 2011). The data obtained from interview were then transcribed and finally analyzed descriptively.

3. Result and Discussion

The collected data were analyzed by using Miles and Huberman qualitative data analysis. It consists of data reduction that refers to the process of selecting, focusing, simplifying the data in written up field notes or transcription. Then the data are displayed in the form of extended text and the last the data are concluded and verified (Miles & Huberman, 1994).

From the transcribed interview it was found that (one respondent) one of teacher's problems in assessing students' writing is when students' writing is not related to the topic given by the teacher.

Respondent 1: Kadang-kadang ada juga siswa yang disuruhnya apa, dijawabnya apa.Masih ada ya satu dua orang yang seperti itu. Kadang terulang lagu-terulang lagi, iya mutar-mutar disitu, nggak pernah sampai ke tujuan. Ada yang begitu juga. Awalnya ngerjain apa, itu yang dibahas lagi.

Teachers also found that some students' writing are unclear and difficult to read.

Respondent 1: Kesusulitannya, ketika bentuknya tulisan anak, sulitnya membaca kalimat mereka. Itu dibutuhkan waktu ekstra.

Respondent 2: Ya,paling kendala untuk menilai writing itu, adalah tulisan. Kenapa ya sekarang anak-anak sekarang itu tulisannya pada jelek. Anak-anak nulisnya pada kemana weh. Aduh, pas saya misalnya memeriksa malam-malam, pas pake pensil lagi udah gk kebaca.

In addition, teachers are lack of time in scoring the students' writing as they have to mark a large number of students.

Respondent 1: Terus terang ya,saya kerepotan. Saya itu kan menangani hampir 500 yang saya ajar. Di satu pihak guru wajib 24 jam, kemudian Bahasa Inggris itu Cuma 2 jam, sehingga untuk mencapai 24 jam, saya harus ngajar 24 kelas. 1 kelas rata-rata 40. 40 siswa dikali 12 kelas ya 480. Jadi untuk mencapai sempurna yang idealis yg diterapkan di rubrik terus terang, jujur tidak bisa tercover. Kecuali jika kita hanya mengajar sekian kelas, siswanya 20. Untuk menggarapnya tidakmaksimal karena jumlah siswa yang terlalu banyak, otomatis tidak bisa idealis, dan tidak bisa benar-benar memantau sejauh mana kemampuan siswanya.

Respondent 2: Terlalu banyak yang harus diperiksa. Kali berapa orang per kelas, per itemnya berapa. Kalo misalkan menilai berapa anak, kapan ngerjainnya. Udah kendalanya di tulisan anak,makan waktu. Makan waktu kan itu mriksanya.

Sometimes students make a very short writing because they are lack of vocabularies.

Respondent 1: Kadang juga siswa menulis hanya sedikit. Tapi walau pun idenya pendek, kalau apa yang diminta sudah tercover, itu sudah dianggap betul. Yang penting dia sudah memenuhi kriteria. Generic structure nya sudah terjawab mulai dari orientation sampai resolution misalnya dalam teks naratif.

Respondent 2: Anak juga kadang menulisnya dikit bangat. Kang-kadang and-and. Kata sambung itu kan banyak bukan and-and aja, then-then aja juga. Penggunaan kata kerja (yang regular-irregular verb). Itu seperti dilupakan anak.jadi klo sudah disuruh ngarang teh, kemana weh kata kerjanya teh. Padahal dimana-mana ngarang vocabulary penting. Klo gk ada vocabulary, mau ngarang apa pun gk bisa,mau ngomong juga gak bisa karena vocabnya terbatas.

Teacher do not have a certain rubric to assess student's writing (they assess only mark according to their holistic, overall and intuitive response to the students' writing).

Respondent 1: Apakah yang mereka tulis sesuai thema, nyambung nggak. Baru lihat isinya (content), baru kemudian dilihat structure nya, baru lihat tanda bacanya, huruf besar dan kecilnya, kesesuaian antar paragraf, itu bisa kita lihat.

Respondent 2: Kalau rubrik sendiri saya tidak ada, overall aja. Kalo rubrik enggak sih ya. saya lihat dari singkatnya dulu. Dari apa susunan kata, keluesan vocabnya, ada yang gak match itu kan. Atau meaningnya gitu, yang advance dengan ini kan bisa, kita bisa melihat.

There are some problems faced by the teachers in assessing students' writing. It was found that one of teacher's problems in assessing students' writing is that students' writing is not related to the topic given by the teacher.It is related to the study conducted by Cresswell (2000) when students focus on language structure rather than focus on ideas or content that make their composition effective. Teachers also found that some students' writing are unclear and difficult to read. It is still found students at this level (grade 1 of senior high school) have difficulty in their handwriting.In line with (Brown, 2010) stated that assessing writing is no simple task as the raters (teachers) need to consider students' writing ability and

what is to test like paragraph construction or logical development of a main idea? Besides teachers have to be aware of text construction, logical development, teachers also have to pay attention on students' writing.

In addition, teachers are lasck of time in scoring the students' writing as they have to mark a large number of students since they have to teach 24 hours in a week (Permendikbud Nomor 62 Tahun 2013 Tentang Sertifikasi Guru dalam Jabatan Dalam Rangka Penataan dan Pemerataan Guru).Besides that, one teacher assess not only students' writing skill, buat also other skills and assess about 500 students. Therefore, teacher will not be able to make an ideal judgement on students' writing.

Other teachers' problem is that students are lack of vocabularies. Sometimes students focus on the grammatical order in the text rather than the content of the text itself. As Graham & Harris in Hooper said that students who are lack of vocabularies tend to write shorter than the students who have more vocabularies. It shows that some are more skillful than others that it may be caused by students' are still too soon to write, students have poor mechanical skill in writing or what teachers ask the students to write does not reflect the students' interest.

Transcribed interview also showed that teachers do not use a spesific criteria or develop the existing scoring scales in assessing the students' writing for some reasons. Firstly, one of the teacher is not familiar with the writing scoring rubrics from some expert in writing. Secondly, it is a time consuming task to do since the teachers have to assess a large number of students. However, scoring criteria play a crucial role in rating students' writing performance. This is particularly in the case of primary trait or analytic scoring methods where assessment are made in relation to each of a number or criteria design represent central features of the language performance under consideration. Scoring criteria channels the way in which raters perceive and evaluate concrete samples of

langauge performance,and finally, come to assign scores to examinees (McNamara, 1996).

Teachers do not have a particular scoring rubric in assessing the students' wriitng even holistic scoring rubric (figure 2). As White in Weigle, 2009 mentioned that holistic scoring has been widely used in assessing writing because of its practicality. The teachers are usually interested in how they can meet writing goals of the course, how they design writing activity that will make students interested in rather than meet students' needs for further writing goals. They usually use score the students' writing by intuition and their general judgement.). In addition, Weigle (2009) proposed that there are three types of rating scalesthat can be used in assessing students' writing both in classroom assessment and large – scale assessment. But, the teachers did not use one of these scales for the reason that they are familiar with these writing rubrics, therefore they assess the students' writing based on their general judgement. Scoring criteria channels the way in which raters perceive and evaluate concrete samples of langauge performance,and finally, come to assign scores to examinees (McNamara, 1996).

4. Conclusion and Remark

From the discussion above it is concluded that teachers found some problems in assessing students' writing. Teachers found that students' writing is not related to the topic given by the teacher. Teacher also found that it is difficult to check the students' handwriting. In addition, teachers are also lack of time in assessing student's writing since one tecaher have to score for about 500 students' writing. Besides that, it also difficult for teachers to assess the students' writing when students come to write a very simple text because students are lack of vocabularies. In addition, teachers do not have a certain or developed scoring rubric to assess students' wriitng that can give judgement on students'writing in the form of final score.

Based on the problems above, it can be recommended that (1) Teacher has to be aware regarding to the problems they find in assessing students' writing, and (2) Teacher needs to have a certain or developed scoring rubric to assess students' writing because it can help them to give judgement on students' writing and students can know how well they do on writing.

References

Brindley, G. 1998.*Describing languagedevelopment? Rating scales and SLA*. Cambridge: Cambridge University Press.

Brown, H. D. 2010. *Language assessment: principal and classroom practices*. New York: Pearson Education.

Cooper, P. L. (1984). *The assessment of writing Ability: A review of research* (GRE Board Research Report No. GREB 82-15R/ETS Research Report (84-12). Princeton, NJ: Educational Testing Service.

Edward, Mark Evan. 2002. *Writing before students: A model for teaching sociological writing*: Teaching Sociology 30 (2): 54 – 59. *creating and teaching writing intensive courses: teaching sociology*. 27 (4), 310 – 323.

Erford, B. T. (2001). *Reliability and validity of the writing essential skill screener – elementaryversion* (WESS – E).Vol (26) No.3,43 – 55. Loyola College: Maryland.

Fraenkel, J., R. *at.al.* (2011). *How to design and evaluate research in education*: McGraw Hill: New York.

Graham, S., & Harris, K. R. (1997). *It can be taught but it does not developnaturally: Myths and realities in writing instruction*. School Psychology Review, 26, 414-415.

Grauerholz, Liz. 1999. *Creating and teaching writing intensive courses: Teachingsociology*.27 (4): 310 – 323.

Hooper, S. R., Montgomery, J., Swartz, C., Reed, M. S., Sandler, A.D., Levine, M. D., et al. (1994). *Measurement of written language expression.* In G. R. Lyon (Ed), *Frames of reference for the assessment of learning disabilities: New views on measurement issues* (375-417). Baltimore: Brookes.

Huot, B. (1990). *Reliability, validity, and holistic scoring: What we know and what we need to know*. College Composition and Communication, 41 (2), 201-203.

Hudd, Suzanne S. And Eric Franklyn Broson. 2007. *Moving forward looking forward: Teaching sociology*. 35 (3): 264 – 273.

Malcom, Nancy L. 2006. Analyzing the news: *Teaching critical thinking skills in a writing intensive social problems course: teaching sociology*. 34 (2): 143-149.

McNamara, T. F. (1996). *Measuring second language performance.* London: Longman.

Miles, M. B., & Huberman A., M. 1994. *Qualitative data analysis*. UnitedStates of America: Sage Production Editor: Pearson Education: New York.

Phipps, S. & Borg, S. (2009).*Exploring Tension Between Teacher's Grammar Teaching Beliefs and Practices*. (37), pp. 380 – 390.

Weigle, S. C. 2002. *Assessing writing*. Cambridge University Press: United Kingdom.