Development of a Rubric to Assess Academic Writing Incorporating Plagiarism Detectors

Salim Razi

Abstract

Similarity reports of plagiarism detectors should be approached with caution as they may not be sufficient to support allegations of plagiarism. This study developed a 50-item rubric to simplify and standardize evaluation of academic papers. In the spring semester of 2011-2012 academic year, 161 freshmen’s papers at the English Language Teaching Department of Çanakkale Onsekiz Mart University, Turkey, were assessed using the rubric. Validity and reliability were established. The results indicated citation as a particularly problematic aspect, and indicated that fairer assessment could be achieved by using the rubric along with plagiarism detectors’ similarity results.

Keywords

academic writing, academic writing scoring rubric, assessing academic papers, plagiarism, plagiarism detector, Turnitin

Writing academic papers is regarded as a complicated task by students and their assessment is also a challenging process for lecturers. Interestingly, the problems in assessing writing are believed to outnumber the solutions (Speck & Jones, 1998). To overcome this, lecturers have referred to a range of theoretical approaches. To achieve a systematic evaluation process, lecturers normally use a scoring rubric that evaluates various discourse and linguistic features along with specific rules of academic writing. However, recent technological advances appear to contribute to a more satisfactory or accurate assessment of academic papers; for example, “Turnitin” claims to prevent plagiarism and aid online grading. Although such efforts deserve recognition, it is still the lecturers themselves who have to score the assignments; therefore, they need to be able to combine reports from plagiarism detectors with their own course aims and outcomes. In other words, their rubric needs to result in accurate assessment through a fair evaluation process (Comer, 2009). Consequently, this study aims at developing a valid and reliable academic writing assessment rubric, also known as a marking scheme or marking guide, to assess EFL (English as a foreign language) teacher candidates’ academic papers by integrating similarity reports retrieved from plagiarism detectors.

In this respect, the researcher developed the “Transparent Academic Writing Rubric” (TAWR), which is a combination of several essential components of academic writing. Although available rubrics include common characteristics, almost none deals with the appropriate use of in-text citation rules in detail. As academic writing heavily depends on incorporating other studies, students should be capable of administering such rules appropriately themselves, as suggested by Hyland (2009). TAWR included 50 items, each carrying 2 points out of 100. The items were grouped in five categories under the subtitles of introduction (8 items), citation (16 items), academic writing (8 items), idea presentation (11 items), and mechanics (7 items). These items together aimed to assess how reader-friendly the texts were with specific emphasis on the accuracy of referencing as an essential component of academic writing (Moore, 2014).

Plagiarism

Plagiarism is defined as “the practice of claiming credit for the words, ideas, and concepts of others” (American Psychological Association [APA], 2010, p. 171). The challenges caused by plagiarism are becoming more important in parallel with developments in Internet technology. In general, plagiarism may occur in any aspect of daily life such as academic studies, computer games, journalism, literature, music, arts, politics, and many more. Unsurprisingly, higher profile plagiarizers receive more attention from the public (Sousa-Silva, 2014). Recently, in the academic context, more lecturers have been complaining about plagiarized...
Assignment submissions by their students and the worldwide plagiarism issue cannot be restricted to any one country, gender, grade, or language proficiency.

In a related study, Sentleng and King (2012) questioned the reasons for plagiarism; their results revealed the Internet as the most probable source of plagiarism and many of the participants in their study had committed some form of plagiarism. Then, considering the worldwide impact of Internet technology, it could be inferred that plagiarism appears to be a nuisance for any lecturer on Earth. Therefore, making effective use of plagiarism detectors seems to be an unavoidable instrument requiring use by most lecturers.

Assessment Rubrics

In relation to the specific importance that assessment has received over the past two decades (Webber, 2012), various rubrics appear to fulfill the needs of writing lecturers, who choose the most appropriate one in accordance with their aims (Becker, 2010/2011). However, the use of rubrics requires care because they bring disadvantages along with advantages (Hamp-Lyons, 2003; Weigle, 2002). An ideal rubric is accepted as one that is developed by the lecturer who uses it (Comer, 2009). The key issue is therefore developing a rubric to meet the expectations of course outcomes. Nevertheless, as Petruzzi (2008) highlighted, writing teachers are human beings entrusted with the aim of “analysing the thinking and reasoning—equally hermeneutic and rhetorical performances—of other human beings” (p. 239).

Comer (2009) warned that in the case of using a shared rubric, lecturers should interact in “moderating sessions” to enable the defining of shared agreements. Nevertheless, Becker (2010/2011) revealed that U.S. universities usually adopted an existing scale and very few of them designed their own rubrics. To conclude, more valid scoring rubrics can be retrieved by integrating actual samples from student-papers through empirical investigations (Turner & Upshur, 2002). That is the basic aim of this study.

Types of Assessment Rubrics

Relevant literature (e.g., Cumming, 1997; East & Young, 2007) refers to three basic assessment rubrics to accomplish performance-based task evaluation, namely analytic, holistic, and primary trait, that are part of the formal evaluation procedure. Becker (2010/2011) explained that analytic scoring requires in-depth analysis of the components of writing such as unity, coherence, flow of ideas, formality level, and so on. In this approach, each component is represented by a weighted score in the rubric. However, the aspects of unity and coherence might require more detailed examination of ideas.

In holistic scoring, raters quickly acknowledge the strengths of a writer rather than scrutinizing drawbacks (Cohen, 1994). Moreover, Hamp-Lyons (1991) introduced another dimension, focused holistic scoring, in which raters relate students’ scores with their expected performance in general writing skills on a variety of proficiency levels. Despite some problems, the ease of practicality makes holistic scoring a popular assessment type. However, analytic rubrics are known to be increasing in reliability (Knoch, 2009), whereas holistic ones are viewed as providing greater validity (White, 1984) because they enable an overall examination. All things considered, analytic rubrics may assist learners to develop better writing skills (Dappen, Isernhagen, & Anderson, 2008) along with encouraging the development of critical thinking subskills (Saxton, Belanger, & Becker, 2012).

The third type of scoring, primary trait scoring, is also known as focused holistic scoring and is considered the least common (Becker, 2010/2011). This is similar to holistic scoring and requires focusing on an individual characteristic of the writing task. It deals with the vital features of particular types of writing: for instance, by considering differences between several types of essays. Cooper (1977) also deals with multiple-trait scoring in which the aim is achieving an overall score via several subscores of various dimensions. Nevertheless, neither primary nor multiple-trait scoring types are fashionable. For example, Becker’s study on different types of rubrics used to assess writing at U.S. universities indicated no use of primary trait rubrics. To summarize, primary trait scoring is equated to holistic scoring whereas multiple-trait scoring can be associated with analytic scoring (Weigle, 2002).

Rubrics can also be classified in accordance with their functions by regarding whether they measure proficiency or achievement to identify the elements to be included in the assessment rubric (Becker, 2010/2011). Proficiency rubrics aim to reveal the level of an individual in the target language by considering general writing skills (Douglas & Chapelle, 1993) whereas achievement rubrics deal with identifying an individual’s progress by examining specific features in the writing curriculum (Hughes, 2002). However, Becker calls attention to the absence of a clear model to assess general writing ability due to the numerous factors that must be considered. This, in turn, results in questioning the validity of rubrics to measure proficiency (see Harsch & Martin, 2012; Huang, 2012; Zainal, 2012, for recent samples).

In relation to this, Fyfe and Vella (2012) investigated the impact of using an assessment rubric as teaching material. Integration of assessment rubrics into the evaluation process may have an enormous impact on several issues such as “creating cooperative approaches with teachers of widely disparate levels of experience, fostering shared learning outcomes that are evaluated consistently, providing timely feedback to students, and integrating technology-enhanced processes with such rubrics can provide for greater flexibility in assessment approaches” (Comer, 2009, p. 2). Subsequently, Comer specifically deals with inter-rater reliability in the use of common assessment rubrics by several teaching staff.
Although the teachers’ experience has an impact on the evaluation process, Comer assumes that such a problem can be resolved by maintaining interaction among teachers.

**Rater Training**

Assessing writing undoubtedly involves subjective evaluation. That is why the scores assigned to student papers are questionable in terms of reflecting the students’ real writing skills (Knoch, 2007) and, unavoidably, raters have an impact on the scores that students attain (Weigle, 2002). The teaching experience of raters is believed to have an enormous impact on the assigned scores. Thus, rating reliability is considered “a cornerstone of sound performance assessment” (Huang, 2008, p. 202). Therefore, to increase the reliability of rubrics, lecturers should plan their assessment procedure carefully before delivering a task.

Although the relevant literature on the necessity of training raters encourages institutions to take precautions, problems related with a subjective scoring procedure remain. This is crucial as it may account for the considerable variance (up to 35%) found in different raters’ scoring of written assignments (Cason & Cason, 1984). To increase inter-rater reliability, the items in rubrics need more detailed explanation. Similarly, Knoch (2007) blamed “the way rating scales are designed” for variances between raters (p. 109). The solution, therefore, might be to invite raters to develop their own rubrics.

**Electronic Scoring and Plagiarism Detectors**

Technological advances can play a vital role in the assessment of written assignments; thus, as a new phenomenon, the implementation of automated essay scoring (AES) has received heightened importance. Studies have mainly aimed at investigating the validity of the AES procedure (James, 2008). The attractiveness of the idea of bypassing human raters by integrating AES systems was rather stimulating; however, initial attempts yielded in non-supportive results to provide evidence on it (e.g., McCurry, 2010; Sandene et al., 2005). The main criticisms of AES focus on its lack of construct validity. For example, Dowell, D’Mello, Mills, and Graesser (2011) recommended considering the impact of topic relevance in the case of AES. However, there have recently been efforts to develop AES by means of computational methods in which interaction between human and AES procedures can be found.

In one study of AES, McNamara, Crossley, and McCarthy (2010) used the automated tool of Coh-Metrix to evaluate student essays in terms of several linguistic features such as cohesion, syntactic complexity, diversity of words, and characteristics of words. In another study, Crossley, Varner, Roscoe, and McNamara (2013) dealt with two Writing Pal (W-Pal) systems namely intelligent tutoring and automated writing evaluation. In their study, students were instructed on writing strategies and received automated feedback. Increasing the use of global cohesion features led the researchers to draw conclusions on the promising impacts of AES systems. In another study, this time Roscoe, Crossley, Snow, Varner, and McNamara (2014) reported on the correlation between computational algorithms and several measures such as writing proficiency and reading comprehension. Although such studies undoubtedly make a significant contribution to the methodology of teaching writing, it should be remembered that scrutinizing AES procedures in depth is outside the aim of the present study. However, the findings of the relevant studies inspire writing teachers with the hope of incorporating AES in a more valid and reliable manner in the near future.

In addition to AES studies, researchers have also investigated the effect of plagiarism detectors such as Turnitin, SafeAssign, and MyDropBox. Their impact has been exaggerated recently in parallel with rapid changes in digital technology that have made plagiarism such a vital contemporary issue, specifically, regarding university assignments (Walker, 2010). The principal idea behind such tools was detecting expressions that did not originally belong to the students. To enable plagiarism detectors to do this, they refer to several databases consisting of webpages, student papers, articles, and books. Several research studies provide evidence for the effectiveness of plagiarism detectors on both preventing and detecting plagiarism (see the Turnitin [2012] report that consists of 39 independently published studies regarding the impact of plagiarism detectors); however, teachers still need to be alerted against the incidents of plagiarized texts that come from the sources non-existent in the databases of plagiarism detectors. In this respect, Kaner and Fiedler (2008) encouraged scholars to submit their texts such as articles and books to the databases of plagiarism detectors with the hope of increasing the benefits of plagiarism detectors.

Despite the popularity of plagiarism detectors, critical issues in the assessment procedure still exist. For example, Brown, Fallon, Lott, Matthews, and Mintie (2007) questioned the reliability of Turnitin similarity reports, which aim to check student-papers’ unoriginal expressions. This saves hours of work for the lecturers (Walker, 2010); however, lecturers should approach such reports with caution as they may not always indicate genuine plagiarism. On their own, plagiarism detectors cannot solve the problem of plagiarism (Carroll, 2009), and detecting genuine academic plagiarism requires a systematic approach (Meuschke & Gipp, 2013). To provide a fair evaluation, students who accidentally plagiarize because of their inadequacy in reporting others’ ideas should be discriminated from those who intentionally do so. Therefore, the final responsibility for detecting plagiarism belongs to the lecturer, as a human considering the students’ intentions, not to a machine (Ellis, 2012). In this respect, the present study aims to fill the gap by developing a rubric to
assess academic writing in a reliable manner with the help of information retrieved from plagiarism detectors.

Method

The researcher developed TAWR (see Appendix) with the hope of taking all aspects of academic writing rules into consideration to enable both an easy and fair marking process. After providing validity and reliability for TAWR, the study aimed at answering the following three research questions:

Research Question 1: In which category of TAWR do students receive lower and higher scores?
Research Question 2: Do students repeating the course receive higher scores in comparison with regular students?
Research Question 3: Do male students plagiarize more than female students?

Setting

The study was conducted in the English Language Teaching (ELT) Department of Çanakkale Onsekiz Mart University (COMU), Turkey, in the spring semester of the 2011-2012 academic year. The ELT department was appropriate for conducting the study because it was expected that the students would develop academic writing skills in a foreign language as part of their education.

Participants

A total of 272 students were enrolled on the Advanced Reading and Writing Skills course. Of these, either as day or evening students, 142 were taking the course for the first time and 130 were repeating it. Because the ELT department is female dominant, female learners (n = 172) outnumbered male learners (n = 100). The participants’ ages were between 18 and 35 with an average of 21 at the time the data were collected.

Students submitted a 3,000-word review paper at the end of the term to pass the course. Although 272 students registered, 82 did not submit their assignments. The explanation might be related with the deterrent impact of Turnitin (see “Findings and Discussion” section). Before marking the written assignments, the researcher of the present study and also the lecturer on the Advanced Reading and Writing Skills course pre-screened them as explained in “Procedures of Data Collection” section. The researcher rejected further evaluation of 29 papers due to extensive use of two types of plagiarism, namely, verbatim and purloining. This is in keeping with Walker’s (2010) justification in which less than 20% plagiarism is considered “moderate” whereas 20% or more plagiarism is regarded as “extensive” (p. 45). Table 1 shows the acceptance and rejection data on submissions.

| Group       | Gender | n  | Accepted | Rejected | Total |
|-------------|--------|----|----------|----------|-------|
| Day         | Female | 26 | 23       | 3        | 37    |
|             | Male   | 11 | 4        | 7        |       |
| Evening     | Female | 34 | 33       | 1        | 51    |
|             | Male   | 17 | 14       | 3        |       |
| Repeat day  | Female | 35 | 35       | 0        | 58    |
|             | Male   | 23 | 17       | 6        |       |
| Repeat evening | Female | 31 | 26       | 5        | 44    |
|             | Male   | 13 | 9        | 4        |       |
| Total       |        | 190| 161      | 29       | 190   |

Instruments

Validity and reliability are assumed to be the most important characteristics of TAWR; therefore, the rubric was examined bearing these features in mind. Investigation began by consulting related experts. First, a professor acting as head of the Foreign Languages Teaching Department at COMU was consulted. In addition, two assistant professors at COMU examined TAWR. To check the applicability of TAWR to languages other than English, an associate professor in the Turkish Language Teaching Department of COMU was also consulted. This was necessary because studies so far have mainly considered the assessment of writing by developing rubrics for English only (East, 2009).

To establish construct validity, Campbell and Fiske’s (1959) approach was administered, where construct validity constitutes two components, namely, convergent and discriminant validity. Bagozzi (1993) indicated that convergent validity deals with the degree of agreement aiming to measure the same concept by means of multiple methods. On the other hand, discriminant validity aims to reveal the discrimination by measuring different concepts. Therefore, convergent validity requires high correlation to measure the same concepts whereas with discriminant validity, high correlations are not expected to measure unique concepts.

Campbell and Fiske’s (1959) approach investigated convergent and discriminant validity by considering four criteria in the multi-trait–multi-method (MTMM) matrix. Their first criterion aims to establish convergent validity by examining monotrait–heteromethod correlations for the same traits via different methods. However, convergent validity on its own does not guarantee construct validity. Then, in the rest of the MTMM matrix, by means of the other three criteria, they deal with discriminant validity to maximize the reliability of the validity measures. The first discriminant validity criterion is the examination of heterotrait–heteromethod coefficients. The expectation is to retrieve higher correlations in measurement of the same traits by means of different methods. This can be observed in the validity diagonal where the value is higher than the coefficients, either in columns or in
rows, in its adjacent heterotrait–heteromethod triangle. The second discriminant validity criterion expects the superiority of the monotrait–heteromethod coefficients over the corresponding heterotrait–monomethod coefficients. Finally, the third discriminant validity criterion predicts similar correlations among heterotrait–monomethod and heterotrait–heteromethod values.

To administer Campbell and Fiske’s (1959) approach, student papers were scored by means of holistic and analytic rubrics in addition to TAWR. To hold to the original MTMM matrix, the number of categories in TAWR was reduced to three by combining related ones to each other. Then, papers were scored through holistic and analytic rubrics and also TAWR by considering three subcategories, namely, flow of ideas, academic writing rules, and mechanics. Table 2 illustrates correlational comparisons between traits by means of three rubrics.

The data in Table 2 were analyzed by means of Campbell and Fiske’s (1959) criteria. First, the data demonstrate that all diagonal values are large and significantly different from zero ($p < .001$) with correlation values ranging from .67 to .91. Thus, this provides evidence of convergent validity. The first discriminant validity criterion compares 36 correlation values and none of these values show a violation of the related criterion. Similar to the previous criterion, again the second discriminant validity criterion entails 36 correlational comparisons. Only one of the comparisons indicates a violation of directionality: $r_{A1,A2} = .68$, and this is not greater than $r_{A2,C2} = .67$, and this is not greater than $r_{A1,A2} = .68$. However, a careful examination indicates that the magnitude of the difference is relatively small (i.e., $\Delta r = .01$). Because there is one violation to this criterion, the proportion of failure is $p = 1 / 36 = .028$. As the significance value is below the chance level of .05, this violation does not spoil discriminant validity. Finally, the third discriminant validity criterion was taken into consideration by comparing the rank order of correlations across triangles. This revealed that the patterns of correlations were the same among the heterotrait–monomethod triangles (i.e., $r_{A1,A2} > r_{A1,A3} > r_{A2,A3}$ and $r_{A1,C2} > r_{A1,C3} > r_{A2,C3}$). Thus, the third discriminant validity criterion was also met. All in all, Campbell and Fiske’s four criteria provide evidence for both convergent and discriminant validity of TAWR.

Reliability involves consistency in test-taker performance by questioning comparability of their performance. Clearly, the acquisition of more information for reliability analysis results in more reliable conclusions. To check reliability of this study, Cronbach’s alpha test was administered to students’ scores and indicated a reliability of .89 for the 50-item TAWR.

Moreover, inter-rater reliability for TAWR was checked. An experienced lecturer was trained how to use TAWR and scored 55 samples selected from 161 student papers. The students’ overall scores from the first assessment were taken into account when selecting these 55 papers. The scores of 161 students were put in descending order and divided into smaller groups constituting 15 papers each, except for the last group that included 11 papers. Thus, 11 groups were constructed and the first five papers from each group were scored by the second rater to enable papers from all ranges to be checked. The second rater’s scores and the researcher’s

---

Table 2. A Synthetic Multitrait–Multimethod Matrix.

| Traits | A1 | A2 | A3 | B1 | B2 | B3 | C1 | C2 | C3 |
|--------|----|----|----|----|----|----|----|----|----|
| Method I (TAWR) |   |    |    |    |    |    |    |    |    |
| Flow | (.95) |   |    |    |    |    |    |    |    |
| Academic rules |    | .68 | (.96) |    |    |    |    |    |    |
| Mechanics |    | .48 | .35 | (.96) |    |    |    |    |    |
| Method I (Analytic) |   |    |    |    |    |    |    |    |    |
| Flow | .91 | .62 | .42 |    |    |    |    |    |    |
| Academic rules |    | .48 | .80 | .20 |    |    |    |    |    |
| Mechanics |    | .34 | .30 | .75 |    |    |    |    |    |
| Method I (Holistic) |   |    |    |    |    |    |    |    |    |
| Flow | .81 | .52 | .46 |    |    |    |    |    |    |
| Academic rules |    | .43 | .67 | .43 |    |    |    |    |    |
| Mechanics |    | .49 | .36 | .77 |    |    |    |    |    |

Source. Adapted from Campbell and Fiske (1959, p. 82).

Note. N = 161. The validity diagonals are the three sets of italicized values. The reliability diagonals are the three sets of values in parentheses. Each heterotrait–monomethod triangle is enclosed by a solid line. Each heterotrait–heteromethod triangle is enclosed by a broken line. TAWR = Transparent Academic Writing Rubric.
scores were positively correlated, Pearson’s $r(55) = .97, p < .001$, with strong evidence for the reliability of TAWR.

Finally, intra-rater reliability was checked. The researcher rescored the same 55 samples following a gap of 6 months after the first assessment in order not to be prejudiced by the first assessment. The first and second scores were positively correlated, Pearson’s $r(55) = .99, p < .001$, providing additional evidence of the reliability of TAWR.

**Procedures of Data Collection**

Students took the Advanced Reading and Writing course for two semesters in which the course contents suggested by Razi (2011) were followed. During 21 hr of classroom instruction in the fall semester, students learned APA in-text citation rules on blending sources. Furthermore, in 21 hr of classroom instruction in the spring semester, they were instructed on reviewing literature, the components of academic papers, headings, writing reference entries, and presenting tables and figures. Thus, the lecturer aimed to provide a positive identity in writing related to the requisites of academic writing as suggested by Lavelle and Zuercher (2001).

Before being marked, the assignments were pre-screened to decide whether each submitted assignment deserved further evaluation or not, which started with a review of student portfolios. During the semester, the students attended 5-min individual tutorial sessions on six occasions and kept records of step-by-step supplementary files such as a transcription of brainstorming on their topic, their assignment outline, the first and second drafts, and revised and proofread versions. Investigation of portfolios and attendance of tutorials gave a general idea regarding their performance. Second, the length of the assignments (requirement was a 3,000-word paper) was considered. The word count for submitted assignments ranged from 1,220 to 5,733 with an average of 2,872. Then, some overall scores were decreased in accordance with the number of words in their papers. As illustrated in Table 3, they were not penalized for the first 300 missing words. It is worth mentioning that the emphasis was not on the length of papers, instead, the quality of academic writing was prioritized; yet, it would not be fair to disregard students who took the easy way out with less effort in terms of length when writing their papers.

In the third step of the pre-screening process, the quotation ratio was retrieved from Turnitin. Although academic writing requires quotations, the ratio is important because novice authors simply quote several expressions by adding them to each other without blending them into their discussion. Such excessive use of quotations should be penalized to encourage paraphrasing. Thus, a ratio of more than 10% was regarded as excessive and the exceeded figure was used as a penalty point.

The final step of pre-screening dealt with plagiarism concerns. It is essential to maintain a consistent evaluation procedure for plagiarizers. Although Turnitin reveals similarities between student papers and other potential sources, the ratio of a similarity report does not necessarily indicate plagiarism. As listed by Glendinning (2014), there is no standard procedure regarding sanctions for plagiarism. A threshold of 20% or more was considered excessive, just above Barrett and Malcolm’s (2006) 15% figure. In the case of detected plagiarism up to 20%, Turnitin similarity reports were used as evidence for a penalty by subtracting the similarity ratio from the paper’s overall score. In doing this, it is important to exclude quotations and reference lists from similarity reports. Nevertheless, because the freshmen in this study were novice authors in academic writing, isolated instances of plagiarism were not considered plagiarism as such when consisting of short portions of copied but not cited expressions in a single sentence. Students whose similarity reports were 20% or higher were informed that they needed to rewrite their papers properly.

**Procedures of Data Analysis**

SPSS 20.0 was used to analyze the data. Descriptive statistics analyzed students’ demographic information and the items in TAWR. Cronbach’s alpha reliability score was established and independent samples $t$ tests were administered to compare gender differences along with differences related to the students’ status, that is, whether they were taking the course for the first time or not. Finally, Pearson’s correlation checked the relation between student scores and several variables.

**Limitations of the Study**

Despite the care taken with the methodology of this study, there are several limitations. The first limitation concerns Turnitin similarity reports. Although Turnitin refers to several databases to identify unoriginal expressions, it is still possible to find sources that do not exist in their databases. Thus, similarity reports may not indicate actual plagiarism ratios (McKeever, 2006; Walker, 2010). The second limitation is that despite indicative results, they may not be

| Number of words | Penalty points |
|-----------------|---------------|
| 2,700 to 3,000  | None          |
| 2,699 to 2,400  | 10            |
| 2,399 to 2,100  | 20            |
| 2,099 to 1,800  | 30            |
| 1,799 to 1,500  | 40            |
| 1,499 to 1,200  | 50            |
| 1,199 to 900    | 60            |
| 899 to 600      | 70            |
| 599 to 300      | 80            |
| Less than 300   | 90            |
Findings and Discussion

Research Question 1

The results indicated that students received their lowest mean scores \((M = 1.32)\) in the subgroup “citation” related to following APA rules for in-text citations and references. The second lowest score \((M = 1.38)\) appeared in the category of “idea presentation” where the flow of ideas was considered. The category of “introduction” dealt with items such as topic selection and presentation, and received an average score \((M = 1.47)\). One of the least problematic groups was “mechanics” \((M = 1.54)\) where grammar, spelling, paper format, and other similar issues were considered. Finally, the characteristics of academic writing received the highest score \((M = 1.64)\).

Figure 1 illustrates students’ mean values in five categories. Apart from the mean values in each category, students received their highest scores for “topic selection,” “match of citations with reference entries,” and “use of tables and figures.” However, their scores on “use of in-text citation rules,” “citing when necessary,” and “ratio of quotes” were very low in terms of the items in TAWR.

Research Question 2

An independent-samples \(t\) test did not indicate significant differences between regular students’ \((M = 34.97, SD = 38.05)\) and repeating students’ \((M = 33.01, SD = 30.78)\), \(t(270) = .47, p = .64\), overall scores. However, this analysis included students whose papers were rejected because of plagiarism along with students who did not submit their assignments. Therefore, Table 4 might be beneficial regarding the total number of submissions by considering the students’ status on the course.

Although the incidents of failure and success were similar to each other in terms of students being either a regular or repeating student, the number of repeating students not submitting their assignment was almost double that of regular students. Hence this does not necessarily mean that repeating students possessed an advantage in terms of being successful on the course.

Despite the assessment of students by means of their final products, evaluation was a continuous process carried out when they submitted brainstormed items, outlines, drafts, revised and proofread versions. This supports student learning (Hernandez, 2012). Such a procedure in a writing class includes basic steps that are essential in writing, as identified almost half a century ago in the influential study of Rohman and Wlecke (1964). Because the lecturer provided feedback in tutorials, the course structure combined formative and summative assessments. Moreover, such process writing integrated reading skills by inviting students to read relevant studies; thus, the assessment procedure benefited in terms of authenticity and validity.

The number of non-attendance hours on the course by regular students and their overall scores were negatively correlated, Pearson’s \(r(88) = -.32, p = .003\). Therefore, it could be implied that non-attendance resulted in lower scores. Repeating students were excluded from this analysis as they were not obliged to attend the lectures regularly as an institutional regulation.

Student attendance of the tutorials and their overall scores were positively correlated, Pearson’s \(r(88) = .48, p < .001\). Therefore, it might be suggested that attending tutorials regularly provided the opportunity to attain higher scores. Again, repeating students were disregarded from this analysis.

In conclusion, although the students repeating the course were exposed to this process far more than the others, their exposure did not contribute to their scores either negatively or positively. The explanation might be related with receiving feedback, because submitting written assignments does not guarantee receiving beneficial feedback for assignments submitted to other lecturers of departmental courses.

Research Question 3

Turnitin similarity reports and students’ overall scores were negatively correlated, Pearson’s \(r(161) = -.51, p < .001\). Therefore, it could be implied that increased similarity reduced the assigned score. Besides this, the number of papers rejected by the researcher without further analysis due to verbatim and purloining indicated that male students \((n = 20)\) exceeded female students \((n = 9)\). Specifically, when the

| Status      | n    | Submitted | Non-submitted | Fail | Succeed |
|-------------|------|-----------|---------------|------|---------|
| Regular     | 142  | 88        | 54            | 85   | 57      |
| Repeating   | 130  | 102       | 28            | 83   | 47      |
| Total       | 272  | 190       | 82            | 168  | 104     |

Table 4. Student Submissions and Success in Terms of Status.
total number of female students \((n = 126)\) and male students \((n = 64)\) are considered, almost one third of male students’ papers \((31.25\%)\) were rejected in comparison with female students \((7.14\%)\). Thus, male students seem more prone to plagiarism than female students as opposed to the findings of a recent study by Khang et al. (2014).

This finding might be important as the relevant literature does not indicate any gender differences on plagiarism (e.g., Walker, 2010). However, parallel with the relevant literature, the findings highlighted that first-year undergraduates may experience problems with plagiarism due to their ignorance in academic writing (e.g., Park, 2003; Yeo & Chien, 2007). For example, in one of the first studies that investigated plagiarism (Karlins, Michaels, & Podlogar, 1988), it was reported that 3% of the students plagiarized by copying the works of previous students. Moreover, lecturers should remember that students might not feel that cheating on assignments is a serious problem (Brent & Atkinson, 2011).

An independent-samples \(t\) test indicated significant differences between male students’ \((M = 25.66, SD = 32.96)\) and female students’ \((M = 38.90, SD = 34.88)\), \(t(270) = 3.08, p = .002, d = 0.39\), overall scores with a moderate effect size. Thus, plagiarism lowered the overall scores attained by male students. It should be remembered that this analysis involved students whose papers were rejected because of plagiarism along with students who did not submit their assignments.

**Conclusion**

Notwithstanding the limitations, TAWR appeared to succeed as a scoring rubric with a high degree of validity and reliability. The following conclusions could be drawn with reference to the results.

First, it can be concluded that “citation” was the most problematic aspect of writing a review paper followed by “idea presentation.” Therefore, writing lecturers should focus more on these two issues in their curriculum.

Second, repeating the course does not automatically seem to result in better scores in academic writing. This implies that although students gain more maturity with the help of other ELT-related courses in the following years, solving the problems they experience with academic writing requires separate skills that can only be gained by regular attendance at lectures and tutorials. Specifically, attending both lectures and tutorials is expected to assist those who have worries about writing review papers.

Third, Turnitin similarity reports clearly have an impact on students’ overall scores in academic writing. This could be supposed to provide evidence on the contribution of Turnitin to a fair evaluation process. However, lecturers need to identify how to approach these reports, such as including or excluding references, small matches, quoted expressions, and so on.

Finally, it could be concluded that male students plagiarize more than female students. This provides evidence that male students perform less well in higher education (e.g., Severiens & ten Dam, 2012). Thus, in terms of verbatim and purloining, male learners may be regarded as more suspicious. To provide a fair assessment, it might be assumed that students who submit their assignments deserve to receive a score greater than “0” because students who do not submit their assignments also receive “0”; therefore, their efforts need to be appreciated. However, plagiarism should be the determining factor in such circumstances. In the case of plagiarized papers, students’ efforts to submit the assignment cannot be taken into consideration because it is not possible to give a score for students’ goodwill in such cases.

**Implications**

The findings support the relevant literature in that university students show a tendency toward plagiarizing (e.g., Sentleng & King, 2012). Yet nothing can be done about the problem unless their reasons for plagiarizing are investigated. Although previous research studies provide reasons, the issues should be continuously monitored as recent technological developments might have changed the reasons for plagiarism. For example, although the studies of Park (2003) and Walker (2010) partly took this issue into consideration, there is a need for more detailed further investigation, specifically in English-for-academic-purposes settings.

More importantly, prevention of plagiarism might only be possible with the cooperation of colleagues; thus, there is a need for institutional policies related to academic integrity. In a relevant study, Bretag et al. (2011) highlighted that “an exemplar policy needs to provide an upfront, consistent message, reiterated throughout the entire policy, which indicates a systemic and sustained commitment to the values of academic integrity and the practices that ensure it” (p. 4). Otherwise, it is clear that individual attempts by lecturers will be fruitless. Therefore, the use of plagiarism detectors should be encouraged by institutions as the participation of each lecturer in the database brings new opportunities to detect student plagiarism. Such an approach should be employed worldwide as plagiarism is a common problem. For instance, there are individual attempts to prevent plagiarism at the institution of the present study’s researcher. As the results of the study acknowledge, such attempts cannot be enough to prevent it.

It is clear that more than a quarter of the students did not submit their assignments due to the possible deterrent impact of Turnitin. Although the collected data in this study did not aim to provide clear evidence on Turnitin having such an impact, the researcher, who has been teaching the Advanced Reading and Writing Skills course for several years at the same institution, hypothesizes that these students preferred not to submit a plagiarized paper as their efforts would be in vain. Further research might check this hypothesis by considering whether plagiarism occurs as a result of pedagogical or moral deficiencies, as questioned by McCulloch (2012).

Many of the exiting rubrics on assessing writing enable holistic evaluation. This might be problematic for two basic
reasons. First, the scores assigned may not reflect actual writing skills but result from variations among raters. Second, students cannot readily receive feedback through holistic scoring. However, TAWR may assist lecturers in adopting Assessment for Learning (AfL), as the theory of AfL focuses on accelerating student learning (Davison & Leung, 2009). Thus, TAWR might also be used as a teaching material in the classroom. Such an application was investigated by Fyfe and Vella (2012). With reference to their speculative results, further research might interest itself in identifying the impact of using TAWR as a teaching tool. As the results of the present study provided high inter- and intra-rater reliability for TAWR, Comer’s (2009) warnings are taken into consideration regarding use of the same rubric by several teaching staff.

The assessment procedure of this study appears to be beneficial as TAWR meets Advanced Reading and Writing Skills course outcomes. In addition, feedback provided by the lecturer throughout the term on the development of student papers is crucial, as Comer (2009) reminded us. However, such feedback should also be provided on the final work of students. Therefore, it might be a good idea to share a copy of TAWR before submission of assignments so that students can arrange their papers accordingly. In addition, TAWR can be used in relation with peer-assisted writing (see Topping, 1996 for more on peer-assisted writing). Moreover, following assessment, lecturers should be encouraged to share a copy of TAWR with every individual student. In this way, students can learn the strong and weak points in their papers; on the other hand, the transparency in the evaluation also encourages lecturers to adopt a more detailed evaluation process. It should be borne in mind that such a detailed analysis of student papers enables diagnostic assessment, which is a relatively new phenomenon.

As discussed by Dahl (2007), Turnitin cannot be regarded only as a plagiarism detector. Instead, students should be encouraged to use Turnitin as a device that assists them throughout their writing process by enabling the submission of drafts and then working on them by considering similarity reports. Acquisition of effective practices is known to be the most essential component of developing academic writing skills (Davis & Carroll, 2009); therefore, further research should consider the impact of Turnitin in the writing process rather than merely focusing on its role in detecting plagiarism.

Criticisms against using plagiarism detection software has resulted in a new movement where lecturers aim to cope with plagiarism without using such detectors (Brown, Jordan, Rubin, & Arome, 2010). In this respect, the use of TAWR might be beneficial for those who wish to avoid using plagiarism detector. Instead, it might be used as classroom teaching material to enable students’ self evaluation.

As the results of the current study indicated, the ratio of success is rather low on the Advanced Reading and Writing Skills course; therefore, students need to be more motivated to write their assignments. As identified by Jung (2013), motivation at university is related to several factors and one way of increasing motivation could be to relate the task to real-life situations. Thus, after assessing students’ papers, the researcher of this study encouraged 68 students to revise their papers with reference to the provided comments and then consider the possibility of publishing them on the Internet. Many of the students reacted to this positively and indeed, some had the opportunity to publish online. Thus, deeper analysis into factors that have an impact on motivating students toward academic writing is essential.

Because the items in TAWR are easy-to-score, this might contribute to validity as it reduces inter-rater variances. Recently, such issues have received greater importance especially at European universities due to attempts to create a coherent European System of Higher Education through the Bologna Declaration. As implementation of the European Credit Transfer System requires identifying clear and consistent course outcomes along with appropriate assessment procedures, writing lecturers should follow similar assessment procedures to maximize inter-rater reliability.

Appendix

Transparent Academic Writing Rubric—TAWR (Source original)

| Parts          | Evaluation | Score: |
|----------------|------------|--------|
| Brainstorming  | None       | Acceptable | Excellent | Yes | No |
| Outline        | None       | Acceptable | Excellent | Yes | No |
| First draft    | None       | Acceptable | Excellent | Yes | No |
| Second draft   | None       | Acceptable | Excellent | Yes | No |
| Review         | None       | Acceptable | Excellent | Yes | No |
| Proofread      | None       | Acceptable | Excellent | Yes | No |

Preliminary decision through pre-screening:
( ) The final version is acceptable for further evaluation.

(continued)
The final version is unacceptable for further evaluation because of
( ) the inappropriate length (word count = ...........).
( ) overused quotations (ratio for quotations = .......%).
( ) plagiarism concerns (.......% similarity report).

| Evaluation criteria                                      | Poor | Acceptable | Excellent |
|----------------------------------------------------------|------|------------|-----------|
| Introduction                                             |      |            |           |
| 1 Topic selection                                        | 0    | 1          | 2         |
| 2 Narrowing down the topic                               | 0    | 1          | 2         |
| 3 Title of the paper                                     | 0    | 1          | 2         |
| 4 Headings and subheadings                               | 0    | 1          | 2         |
| 5 Abstract                                                | 0    | 1          | 2         |
| 6 Key words                                               | 0    | 1          | 2         |
| 7 Introduction to the topic                               | 0    | 1          | 2         |
| 8 Mentioning the aims in the introduction                | 0    | 1          | 2         |
| Citation                                                 |      |            |           |
| 9 Citing when necessary                                   | 0    | 1          | 2         |
| 10 Introducing paraphrases and summaries                  | 0    | 1          | 2         |
| 11 Restructuring in paraphrases/summaries                 | 0    | 1          | 2         |
| 12 Rewording in paraphrases/summaries                     | 0    | 1          | 2         |
| 13 Introducing quotes                                     | 0    | 1          | 2         |
| 14 Use of quotations                                      | 0    | 1          | 2         |
| 15 Citing quotes appropriately                            | 0    | 1          | 2         |
| 16 Ratio of quotes                                        | 0    | 1          | 2         |
| 17 Sufficiency of the number of cited sources             | 0    | 1          | 2         |
| 18 Reliability of the cited sources                       | 0    | 1          | 2         |
| 19 Appropriate use of secondary sources                   | 0    | 1          | 2         |
| 20 Ratio of secondary source use                          | 0    | 1          | 2         |
| 21 Appropriate use of in-text citation rules              | 0    | 1          | 2         |
| 22 Writing reference entries                              | 0    | 1          | 2         |
| 23 Order of reference entries                             | 0    | 1          | 2         |
| 24 Exact match of citations with reference entries        | 0    | 1          | 2         |
| Academic Writing                                          |      |            |           |
| 25 Focusing on the issue (omitting personal pronouns)     | 0    | 1          | 2         |
| 26 Appropriate use of abbreviations                       | 0    | 1          | 2         |
| 27 Avoiding contractions                                  | 0    | 1          | 2         |
| 28 Avoiding extremeness                                  | 0    | 1          | 2         |
| 29 Avoiding slang, jargon and clichés                     | 0    | 1          | 2         |
| 30 Use of words with precise meaning                      | 0    | 1          | 2         |
| 31 Use of objective language                             | 0    | 1          | 2         |
| 32 Balanced use of passive forms                          | 0    | 1          | 2         |
| Idea Presentation                                         |      |            |           |
| 33 Appropriate use of markers                             | 0    | 1          | 2         |
| 34 Appropriate use of linking devices                     | 0    | 1          | 2         |
| 35 Flow of ideas                                          | 0    | 1          | 2         |
| 36 Paragraph unity                                        | 0    | 1          | 2         |
| 37 Overall unity                                          | 0    | 1          | 2         |
| 38 Paragraph coherence                                   | 0    | 1          | 2         |
| 39 Overall coherence                                     | 0    | 1          | 2         |
| 40 Appropriate length of paragraphs                       | 0    | 1          | 2         |
| 41 Complexity of the sentences                           | 0    | 1          | 2         |
| 42 Relevance of conclusions with the discussion           | 0    | 1          | 2         |
| 43 Drawing effective conclusions                          | 0    | 1          | 2         |
Appendix (continued)

| Parts | Evaluation | Tutorial att. |
|-------|------------|---------------|
| Mechanics | | |
| 44 Paper format | 0 | 1 | 2 |
| 45 Grammar | 0 | 1 | 2 |
| 46 Spelling | 0 | 1 | 2 |
| 47 Punctuation | 0 | 1 | 2 |
| 48 Vocabulary selection | 0 | 1 | 2 |
| 49 Use of tables and figures | 0 | 1 | 2 |
| 50 Length of the paper | 0 | 1 | 2 |

Additional comments:

Author's Note

The author of the present study is an independent researcher and has no relationship whatsoever with the proprietors of Turnitin software other than as a user of the product. The researcher does not in any way intend to promote the commercial interests of Turnitin over any other detection software, but simply presents the experiences of a lecturer in using the tool to detect and prevent plagiarism.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research and/or authorship of this article.

References

American Psychological Association. (2010). Publication manual of the American Psychological Association (6th ed.). Washington, DC: Author.

Bagozzi, R. P. (1993). Assessing construct validity in personality research: Applications to measure of self-esteem. Journal of Research in Personality, 27, 49-87.

Barrett, R., & Malcolm, J. (2006). Embedding plagiarism education in the assessment process. International Journal for Academic Integrity, 2(1), 38-45.

Becker, A. (2010/2011). Examining rubrics used to measure writing performance in U.S. intensive English programs. The CATESOL Journal, 22(1), 113-130. Retrieved from http://www.catesol.org/Becker%20113-130.pdf

Brent, E., & Atkinson, C. (2011). Accounting for cheating: An evolving theory and emergent themes. Research in Higher Education, 52, 640-658.

Bretag, T., Mahmud, S., Wallace, M., Walker, R., James, C., Green, M., . . . Partridge, L. (2011). Core elements of exemplary academic integrity policy in Australian higher education. International Journal for Educational Integrity, 7(2), 3-12.

Brown, R., Fallon, B., Lott, J., Matthews, E., & Mintie, E. (2007). Taking on Turnitin: Tutors advocating change. The Writing Center Journal, 27, 7-28.

Brown, V., Jordan, R., Rubin, N., & Arome, G. (2010). Strengths and weaknesses of plagiarism detection software. Journal of Literacy and Technology, 11, 110-131.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. Psychological Bulletin, 56, 81-105.

Carroll, J. (2009). Should we use Turnitin at this university? Journal of Hospitality, Leisure, Sport & Tourism Education, 8, 157-166.

Cason, G. J., & Cason, C. L. (1984). A deterministic theory of clinical performance rating. Evaluation & the Health Professions, 7, 221-247.

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

Comer, K. (2009). Developing valid and reliable rubrics for writing assessment: Research and practice. Retrieved from https://akoacetarao.ac.nz ako-hub/good-practice-publication-grants e-book/resources/pages/developing-valid-and-reliable-rubrics-writing-assessment-research

Cooper, C. R. (1977). Holistic evaluation of writing. In C. R. Cooper & L. Odell (Eds.), Evaluating writing: Describing, measuring, judging (pp. 3-31). Urbana, IL: National Council of Teachers of English.

Crossley, S. A., Varner, L. K., Roscoe, R. D., & McNamara, D. S. (2013). Using automated indices of cohesion to evaluate an intelligent tutoring system and an automated writing evaluation system. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), Artificial intelligence in education (pp. 269-278). Berlin, Germany: Springer.

Cumming, A. (1997). The testing of writing in a second language. In C. Clapham & D. Corson (Eds.), Encyclopedia of language and education: Language testing and assessment (Vol. 7, pp. 131-139). Norwell, MA: Kluwer Academic.

Dahl, S. (2007). Turnitin®: The student perspective on using plagiarism detection software. Active Learning in Higher Education, 8, 173-191.
Dappen, L., Isernhagen, J., & Anderson, S. (2008). A statewide writing assessment model: Student proficiency and future implications. *Assessing Writing, 13*, 45-60.

Davis, M., & Carroll, J. (2009). Formative feedback within plagiarism education: Is there a role for text-matching software? *International Journal for Educational Integrity, 3*(2), 58-70.

Davison, C., & Leung, C. (2009). Current issues in English language teacher-based assessment. *TESOL Quarterly, 43*, 393-415.

Douglas, D., & Chapelle, C. (1993). Foundations and directions for a new decade of language testing. In D. Douglas & C. Chapelle (Eds.), *A new decade of language testing research* (pp. 1-24). Alexandria, VA: Teachers of English to Speakers of Other Languages.

Dowell, N., D'Mello, S. K., Mills, C., & Graesser, A. (2011). Does topic matter? Topic influences on linguistic and rubric-based evaluation of writing. In G. Biswas, S. Bull, J. Kay, & A. Mitrovic (Eds.), *Proceedings of the 15th international conference on artificial intelligence in education* (pp. 435-437). Berlin, Germany: Springer.

East, M. (2009). Evaluating the reliability of a detailed analytic scoring rubric for foreign language writing. *Assessing Writing, 14*, 88-115.

East, M., & Young, D. (2007). Scoring L2 writing samples: Exploring the relative effectiveness of two different diagnostic methods. *New Zealand Studies in Applied Linguistics, 13*, 1-21.

Ellis, C. (2012). Streamlining plagiarism detection: The role of electronic assessment management. *International Journal for Educational Integrity, 8*(2), 46-56.

Ely, B., & Vella, C. (2012). Assessment rubric as teaching tool: Learning how to “tick all the boxes.” *Cambridge ESOL: Research Notes, 48*(2), 30-36. Retrieved from http://www.cambridgeenglish.org/images/23165-research-notes-48.pdf

Glendinning, I. (2014). Responses to student plagiarism in higher education across Europe. *International Journal for Educational Integrity, 10*(1), 4-20.

Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In L. Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts* (pp. 241-276). Norwood, NJ: Ablex.

Hamp-Lyons, L. (2003). Writing teachers as assessors of writing. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 18-89). Cambridge, UK: Cambridge University Press.

Harsch, C., & Martin, G. (2012). Adapting CEF-descriptors for rating purposes: Validation by a combined rater training scale revision approach. *Assessing Writing, 17*, 228-250.

Hernandez, R. (2012). Does continuous assessment in higher education support student learning? *Higher Education, 64*, 489-502.

Huang, J. (2008). How accurate are ESL students’ holistic writing scores on large-scale assessments?—A generalizability theory approach. *Assessing Writing, 13*, 201-218.

Huang, J. (2012). Using generalizability theory to examine the accuracy and validity of large-scale ESL writing assessment. *Assessing Writing, 17*, 123-139.

Hughes, A. (2002). *Testing for language teachers*. Cambridge, UK: Cambridge University Press.

Hyland, T. A. (2009). Drawing a line in the sand: Identifying the borderzone between self and other in EL1 and EL2 citation practices. *Assessing Writing, 14*, 62-74.

James, C. L. (2008). Electronic scoring of essays: Does topic matter? *Assessing Writing, 13*, 80-92.

Jung, J. Y. (2013). Amotivation and indecision in the decision-making processes associated with university entry. *Research in Higher Education, 54*, 115-136.

Kaner, C., & Fiedler, R. L. (2008). A cautionary note on checking software engineering papers for plagiarism. *IEEE Transactions on Education, 51*, 184-188.

Karlns, M., Michaels, C., & Podlogar, S. (1988). An empirical investigation of actual cheating in a large sample of undergraduates. *Research in Higher Education, 29*, 359-364.

Khang, D. B., Ba, K. D., Quoc, D. L., Binh, A. L. D. T., Phuong, L. N., Phuong, Q. N., & Quoc, L. P. (2014, May 16-18). *Student plagiarism in higher education in Vietnam: An exploratory study*. Paper presented at the 6th International Integrity and Plagiarism Conference: Promoting authentic assessment, Newcastle upon Tyne, UK.

Knoch, U. (2007). “Little coherence, considerable strain for reader”: A comparison between two rating scales for the assessment of coherence. *Assessing Writing, 12*, 108-128.

Knoch, U. (2009). Diagnostic assessment of writing: A comparison of two rating scales. *Language Testing, 26*, 275-304.

Lavelle, E., & Zuercher, N. (2001). The writing approaches of university students. *Higher Education, 42*, 373-391.

McCulloch, S. (2012). Citations in search of a purpose: Source use and authorial voice in L2 student writing. *International Journal for Educational Integrity, 8*(1), 55-69.

McCurry, D. (2010). Can machine scoring deal with broad and open writing tests as well as human readers? *Assessing Writing, 15*, 118-129.

McKeever, L. (2006). Online plagiarism detection services—Saviour or scourge? *Assessment & Evaluation in Higher Education, 31*, 155-165.

McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication, 27*, 57-86.

Meuschke, N., & Gipp, B. (2013). State-of-the-art in detecting academic plagiarism. *International Journal for Educational Integrity, 9*(1), 50-71.

Moore, E. (2014). Accuracy of referencing and patterns of plagiarism in electronically published theses. *International Journal for Educational Integrity, 10*(1), 42-55.

Park, C. (2003). In other (people’s) words: Plagiarism by university students—Literature and lessons. *Assessment & Evaluation in Higher Education, 28*, 471-488.

Petruzzii, A. (2008). Articulating a hermeneutic theory of writing assessment. *Assessing Writing, 13*, 219-242.

Razi, S. (2011). *Advanced reading and writing skills in ELT: APA style handbook*. Ankara, Turkey: Nobel.

Rohman, D. G., & Wlecke, A. O. (1964). *Pre-writing, the construction and application of models for concept formation in writing* (East Lansing. Report #: CRP-2174, ERIC Document Reproduction Service No. ED 001273). East Lansing: Michigan State University.

Rosclo, R. D., Crossley, S. A., Snow, E. L., Varner, L. K., & McNamara, D. S. (2014). Writing quality, knowledge, and comprehension correlates of human and automated essay scoring. In W. Eberle & C. Boonthum-Denecke (Eds.), *Proceedings of the twenty-seventh international Florida Artificial Intelligence...*
Salim Razi works at the ELT Department of Çanakkale Onsekiz Mart University, Turkey. His main research interests include schema theory, reading comprehension, assessing reading and writing, metacognitive reading strategies, academic writing, plagiarism in academic writing, and teaching culture.