Abstract
Concerns with plagiarism are an important barrier in assessment of academic papers. Although Turnitin appears to facilitate the detection of plagiarism to some extent, similarity reports should be approached with caution as they may not be sufficient to support allegations of plagiarism. The present study developed a 50-item rubric to simplify and standardize evaluation that also takes Turnitin results into consideration. At the end of the spring semester of 2011-2012 academic year, 161 freshmen’s papers for the Advanced Reading and Writing Course at the Department of English Language Teaching of Çanakkale Onsekiz Mart University, Turkey were assessed using the rubric. Validity and reliability were established for the rubric. The results indicated citation as a problematic aspect along with gender differences related with plagiarism concerns. It can be inferred that a fairer assessment procedure might be achieved by using the rubric along with Turnitin similarity results.

Key words: academic writing, assessing academic papers, academic writing scoring rubric, plagiarism, plagiarism detector, Turnitin
Assessing academic writing: Development of a rubric and relating Turnitin reports

Assist. Prof. Dr. Salim Rază
Çanakkale Onsekiz Mart University
English Language Teaching Department
salimrazi@gmail.com

Introduction

- The problems in assessing writing are more than just solutions (Spat & Jurus, 1998).
- Recent technological advances:
  - E.g., Turnitin claims to prevent plagiarism and aid online grading.
  - It is still the lecturer who has to score the assignments.
- This study aims at:
  - developing a valid and reliable academic writing assessment rubric
  - by combining similarity reports retrieved from plagiarism detectors.

Types of assessment rubrics

- Analytic scoring rubrics:
  - In-depth analysis of the components such as unity, coherence, flow of ideas, formality level, etc.
  - increase reliability (Novak, 2003).
  - develop better writing skills (Dreyen, Tumul, & Anderson, 2013).
  - development of critical thinking sub-skills (Saxton, Belanger, & Becker, 2012).
- Holistic scoring rubrics:
  - quickly acknowledge the strengths of a writer rather than scrutinizing drawbacks (Cohen, 1994).
  - the ease of practicality; a popular assessment type.
  - greater validity (Whit, 1984) enable an overall examination.

Primary trait vs. Multiple trait scoring

- Primary trait scoring:
  - deals with the vital features of particular types of writing; for instance, by considering differences among several types of essays.
  - the least common (Becker, 2010/2011).
  - equated to holistic scoring
- Multiple trait scoring:
  - achieving an overall score via several subscores of various dimensions (Cooper, 1977).
  - associated with analytic scoring (Weigle, 2002).

Assessment rubrics

- Consider advantages and disadvantages (Hamp-Lyons, 2003; Weigle, 2002).
- A rubric should be developed by the lecturer who uses it (Comer, 2009).
- A rubric should meet the expectations of course outcomes.
- Moderating sessions: Interact to enable defining shared agreements (Comer, 2009).
- Very few universities design their own rubrics (Becker, 2010/2011).
- Integrate actual samples from student papers through empirical investigations (Turner & Upshur, 2002).
Use of assessment rubrics

- Using an assessment rubric as teaching material (Fyfe & Vella, 2012).
- Inter-rater reliability in the use of common assessment rubrics (Comer, 2009).
- Teachers’ experiences have an impact on the evaluation process.
- Comer: Such a problem can be resolved by maintaining interaction among teachers.

Electronic scoring

- Validity of the machine-markable scoring procedure (James, 2008).
- Bypassing human raters???
  - non-supportive results of relevant studies (e.g., McCurry, 2010; Sandene, et al. 2009).
- Researchers are soon expected to develop computer-raters that can score as reliably as human raters.

Problems with plagiarism detectors

- Saves hours of work for the tutors (Walker, 2010)
- Reliability of Turnitin similarity reports (Brown et al., 2007).
- To provide a fair evaluation:
  - Discriminate students who accidentally plagiarise because of their inadequacy in reporting others’ ideas should be from those who intentionally do so.
  - The final responsibility belongs to the tutor.
- A well-developed rubric may provide a desirable impact with the support of Turnitin similarity reports.

Rater training

- Subjective evaluation of writing.
- Scores are questionable; students’ real writing skills (Knoch, 2007).
- Raters’ teaching experiences have an impact on the scores that students attain (Weigle, 2002).
- To increase the reliability of rubrics:
  - plan assessment procedure carefully before delivering the task.

Plagiarism detectors

- Electronic scoring is problematic but not online plagiarism detectors.
- Internet technology makes plagiarism a crucial problem:
  - specifically for university assignments (Walker, 2010).
  - Detect expressions that do not originally belong to students.
- Refers to several databases:
  - web pages,
  - student papers,
  - articles, and
  - books.

The study
Conducted in the English Language Teaching Department of Çanakkale Onsekiz Mart University, Turkey:
In the spring semester of the 2011-2012 academic year.
Appropriate to conduct the study with reference to the proficiency in English of the participants.

Acceptance and Rejection Data on the Students’ Written Assignments

<table>
<thead>
<tr>
<th>Groups</th>
<th>Gender</th>
<th>N</th>
<th>Accepted</th>
<th>Rejected</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day</td>
<td>Female</td>
<td>26</td>
<td>23</td>
<td>3</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>11</td>
<td>4</td>
<td>7</td>
<td>18</td>
</tr>
<tr>
<td>Evening</td>
<td>Female</td>
<td>34</td>
<td>33</td>
<td>1</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>17</td>
<td>14</td>
<td>3</td>
<td>34</td>
</tr>
<tr>
<td>Repeat Day</td>
<td>Female</td>
<td>35</td>
<td>35</td>
<td>0</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>23</td>
<td>17</td>
<td>6</td>
<td>46</td>
</tr>
<tr>
<td>Repeat Evening</td>
<td>Female</td>
<td>31</td>
<td>26</td>
<td>5</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>Male</td>
<td>13</td>
<td>9</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>190</td>
<td>161</td>
<td>29</td>
<td>190</td>
</tr>
</tbody>
</table>

Reviewers’ overall conclusion:
Highlighted the validity of TAWR.

Construct validity:
A professor, head of the Foreign Languages Teaching Department at ÇOMU was consulted.
Two assistant professors in the ÇOMU examined TAWR.
Applicability of TAWR with languages other than English:
An associate professor in the Turkish Language Teaching Department of ÇOMU was consulted.
Relevant studies considered assessing writing by developing rubrics related to English only (East, 2009).

Instrument Validity

Cronbach alpha .89 for the 50-item TAWR.
Inter-rater reliability:
An experienced instructor was trained on how to use TAWR.
Scored 55 samples selected from 161 student papers.
Students’ overall scores from the first assessment were considered.
Put into descending order and divided into smaller groups constituting of 15 papers.
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.
Pearson’s r(55) = .97, p < .001

Intra-rater reliability:
The researcher rescored the same 55 samples with a gap of 6 months.
Pearson’s r(55) = .99, p < .001
ITEMS IN TAWR

TAWR: combination of several essential components of academic writing.

None deals with the appropriate use of in-text citation rules.

50 items

Each 2 points out of 100.

Five categories of items:
- introduction (8 items),
- citation (16 items),
- academic writing (8 items),
- idea presentation (11 items),
- mechanics (7 items).

Procedures of data collection

1st step of the pre-screening

Investigation of student portfolios:
- Decide for further evaluation.
- Students attended five-minute individual tutorial sessions on six occasions.
- Kept records of step-by-step supplementary files:
  - 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 about their performance.

2nd step of the pre-screening

The length of the assignments was considered.

The word-count:
- ranged from 1220 to 5733 with an average of 2872.

Overall scores were decreased in accordance with the number of words in their papers.

3rd step of the pre-screening

The quotation ratio was retrieved from Turnitin.

Novice authors simply quote several expressions by adding them to each other without blending them into their discussion.

Penalize excessive use of quotations.

A ratio over 10% was regarded as excessive.

Penalty Points for Immature Papers

<table>
<thead>
<tr>
<th>Number of words</th>
<th>Penalty points</th>
</tr>
</thead>
<tbody>
<tr>
<td>1798-2000</td>
<td>0</td>
</tr>
<tr>
<td>2099-2400</td>
<td>10</td>
</tr>
<tr>
<td>2599-3100</td>
<td>20</td>
</tr>
<tr>
<td>3099-4000</td>
<td>30</td>
</tr>
<tr>
<td>3599-5000</td>
<td>40</td>
</tr>
<tr>
<td>4099-6000</td>
<td>50</td>
</tr>
<tr>
<td>4599-8000</td>
<td>60</td>
</tr>
<tr>
<td>5099-10000</td>
<td>70</td>
</tr>
<tr>
<td>5599-15000</td>
<td>80</td>
</tr>
<tr>
<td>6099-20000</td>
<td>90</td>
</tr>
<tr>
<td>more than 25,000</td>
<td>100</td>
</tr>
</tbody>
</table>
Plagiarism concerns:
- Similarity report does not necessarily indicate plagiarism.
- In the case of detected plagiarism, Turnitin similarity reports were used as evidence for a penalty by subtracting the similarity ratio from the paper’s overall score.
- 1st year students 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.

Procedures of data analysis:
- SPSS 20.0 was used.
- Descriptive statistics: demographic information and the items in TAWR.
- Cronbach’s alpha reliability.
- Independent samples t-tests: gender differences, students’ status either taking the course for the first time or not.
- Pearson correlation: student scores and several variables.

Limitations of the study:
- The first limitation concerns Turnitin similarity reports:
  - It is possible to find sources which do not exist in their databases.
  - Similarity reports may not indicate actual plagiarism ratios (McKeever, 2006; Walker, 2010).
- The second limitation:
  - Despite indicative results, they may not be generalizable as data come from a single university in the Turkish tertiary context.

Findings and discussion:
RQ1: In which category of TAWR do students receive lower scores?
- Apart from the mean values in each category:
  - Highest scores for topic selection
  - Match of citations with reference entries
  - Use of tables and figures
- Lowest scores for:
  - Use of in-text citation rules
  - Citing when necessary
  - Ratio of quotes

RQ2: Do students repeating the course receive higher scores in comparison to regular students?
- An independent samples t-test:
  - No 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.
- This analysis included students whose papers were rejected because of plagiarism along with students who did not submit their assignments.
RQ2: Students’ Submissions and Success in Terms of Regular/Repeating Status

<table>
<thead>
<tr>
<th>Status</th>
<th>N</th>
<th>Submitted</th>
<th>Non-submitted</th>
<th>Fail</th>
<th>Succeed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regular</td>
<td>142</td>
<td>88</td>
<td>54</td>
<td>85</td>
<td>57</td>
</tr>
<tr>
<td>Repeating</td>
<td>130</td>
<td>102</td>
<td>28</td>
<td>83</td>
<td>47</td>
</tr>
<tr>
<td>Total</td>
<td>272</td>
<td>190</td>
<td>82</td>
<td>168</td>
<td>104</td>
</tr>
</tbody>
</table>

- The incidents of failure and success were similar to each other in terms of being either a regular or repeating student.
- Non-submission was almost doubled that of regular students.
- This does not necessarily mean that repeating students possessed an advantage in terms of being successful on the course.

RQ2: Number of non-attendance hours to the course and overall scores

Pearson correlation analysis:
- significant negative weak and low correlation
  \[ r(88) = -0.32, p = .003 \]
- Non-attendance of the course resulted in lower scores for 10% of the students.
- Repeating students were excluded as they did not have to attend the lectures.

RQ2: Number of attendances to the tutorials and overall scores

Pearson correlation analysis:
- significant weak and low correlation
  \[ r(88) = 0.48, p < .001 \]
- Attending the tutorials regularly provided the opportunity of getting higher scores for 23% of the students.
- Repeating students were disregarded.

RQ3: Do male students plagiarise more than female ones?

<table>
<thead>
<tr>
<th>Plagiarism</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>30</td>
<td>9</td>
</tr>
<tr>
<td>No</td>
<td>176</td>
<td>64</td>
</tr>
<tr>
<td>Total</td>
<td>206</td>
<td>73</td>
</tr>
</tbody>
</table>

Pearson correlation analysis between Turnitin reports and overall scores:
- significant negative moderate correlation \[ r(161) = -0.71, p < .001, r^2 = 0.50 \]
- The increased amount of similarity reduced the assigned score.
- Thus, male students seem more prone to plagiarism than female students.
**RQ3: Do male students plagiarise more than female ones?**

- No gender differences on plagiarism in literature (e.g., Walker, 2010).

**Parallel with literature:**
- First-year undergraduates’ problems with plagiarism due to their ignorance in academic writing (e.g., Park, 2003; Yeo & Chien, 2007).
- E.g., Karlins et al. (1988): one of the first to investigate plagiarism.
- Revealed that 3% of the students plagiarised by copying the works of previous students.
- Students might not feel that cheating on assignments is a serious problem (Krent & Atkinson, 2011).

**Independent samples t-test on overall scores:**
- Significant differences
  - Male (M = 25.66, SD = 32.96)
  - Female (M = 38.90, SD = 34.88)
- t(270) = 3.08, p = .002, d = 0.39
- A moderate effect size.

**Plagiarism lowers male students’ overall scores.**
- This analysis involved students whose papers were rejected because of plagiarism along with students who did not submit their assignments.

**Conclusions**

- Notwithstanding the limitations,
- TAWR appeared to succeed as a scoring rubric with a high degree of validity and reliability.

**Conclusion 1**

- ‘Citation’ was the most problematic aspect of writing a review paper followed by ‘idea presentation’.
- Lecturers should focus more on these two issues in their curriculum.

**Conclusion 2**

- Repeating the course does not automatically result in receiving better scores.
- Students gain more maturity with the help of other ELT related courses in the second year.
- Academic writing requires separate skills that can only be gained by regular attendance at lectures and tutorials.

**Conclusion 3**

- Turnitin similarity reports have an impact on students’ overall scores when writing academic papers.
- Provides evidence on the contribution of Turnitin to a fair evaluation process.
- Identify how to approach these reports:
  - Including or excluding
  - References,
  - Small matches,
  - Quoted expressions, etc.
Male students plagiarise more than females.

Evidence to male students less well performance in higher education (e.g., Severiens & ten Dam, 2012).

Male learners more suspicious in terms of verbatim and purloining.

Fair assessment:

Students who submit their assignments deserve to receive a score greater than 0??

Consider students’ goodwill.

Implication:

Prevention of plagiarism

Prevention of plagiarism requires cooperation of colleagues.

Institutions should take precautions and encourage lecturers to follow them.

Individual attempts would be fruitless.

Encourage plagiarism detectors:

the participation of each lecturer into the database brings new opportunities to detect student plagiarism.

Universities should develop their own campus-wide writing programme assessment (Good et al., 2012).

Implication:

University students show a tendency towards plagiarising.

Investigate reasons for plagiarising.

Previous research studies provide reasons:

Monitor continuously as technological developments might have changed the reasons for plagiarism.

Implication:

Deterrent impact of Turnitin

More than a quarter of the students did not submit their assignments.

Hypothesis for further research:

These students preferred not to submit a plagiarised paper as their efforts would be in vain.

Implication:

TAWR as a teaching material

Exiting rubrics: holistic evaluation.

This might be problematic for two basic reasons.

Variations among raters: Scores may not reflect actual writing skills.

Students cannot receive feedback through holistic scoring.

TAWR may assist lecturers to adopt assessment for learning to accelerate student learning (Davison & Leung, 2009).

Further research:

The effectiveness of using TAWR as a teaching material.

Similar application by Fyfe and Vella (2012).

Implication:

Transparency in providing feedback

Provide feedback throughout the term on the development of student papers (Comer, 2009).

Share a copy of TAWR before submission.

Provide feedback through TAWR on the final works.

Students can learn the strong and weak points in their papers.

Diagnostic assessment: detailed analysis of papers.

The transparency in the evaluation also encourages lecturers to adopt a more detailed evaluation process.
Further research: investigate the impact of Turnitin in the writing process.

68 students were encouraged to publish.

Relate the task to real-life situations.

Students need to be more motivated to write their assignments.

Further research:
- Deeper analysis into motivational factors.

The ratio of success is rather low in the Advanced Reading and Writing Course.

Students need to be more motivated to write their assignments.

The criticism against using plagiarism detection software:
- Lecturers aim to cope with plagiarism without using such detectors (Brown, Jordan, Rubin, & Arome, 2010).
- TAWR might be beneficial for those:
  - as it might be used as classroom teaching material instead to enable student evaluation of their own papers.

Implication: Avoiding Turnitin

The criticism against using plagiarism detection software:
- Lecturers aim to cope with plagiarism without using such detectors (Brown, Jordan, Rubin, & Arome, 2010).
- TAWR might be beneficial for those:
  - as it might be used as classroom teaching material instead to enable student evaluation of their own papers.

Implication: Bologna Declaration

ECTS requires:
- European universities aim to create a coherent European System of Higher Education through the Bologna Declaration.
- ECTS requires:
  - identifying clear and consistent course outcomes and appropriate assessment procedures.
  - Follow similar assessment procedures to maximize inter-rater reliability.

References


Turnitin: by Salim Razi

ICONIE 2013 + Academic writing rubric and Turnitin by Salim Razi

Turnitin & by Salim Razi

ICONIE 2013 & + Academic writing rubric and Turnitin by Salim Razi

Turnitin: Avoiding Turnitin

Further research: investigate the impact of Turnitin in the writing process.

The criticism against using plagiarism detection software:
- Lecturers aim to cope with plagiarism without using such detectors (Brown, Jordan, Rubin, & Arome, 2010).
- TAWR might be beneficial for those:
  - as it might be used as classroom teaching material instead to enable student evaluation of their own papers.

Implication: Motivation

The ratio of success is rather low in the Advanced Reading and Writing Course.

Students need to be more motivated to write their assignments.

The criticism against using plagiarism detection software:
- Lecturers aim to cope with plagiarism without using such detectors (Brown, Jordan, Rubin, & Arome, 2010).
- TAWR might be beneficial for those:
  - as it might be used as classroom teaching material instead to enable student evaluation of their own papers.

Implication: Bologna Declaration

ECTS requires:
- European universities aim to create a coherent European System of Higher Education through the Bologna Declaration.
- ECTS requires:
  - identifying clear and consistent course outcomes and appropriate assessment procedures.
  - Follow similar assessment procedures to maximize inter-rater reliability.

References


References (continue)


Paragraph coherence
Flow of ideas
Appropriate use of linking devices (e.g., however)
Exact match of citations with reference entries
Appropriate use of in-text citation rules
Citing quotes appropriately
Restructuring in paraphrases/summaries
Introducing paraphrases and summaries (variations in the style)

Evaluation criteria
Point(s)

INTRODUCTION
1 Topic selection
2 Narrowing down the topic
3 Title of the paper
4 Headings and subheadings
5 Abstract
6 Key words
7 Introduction to the topic
8 Mentioning the aims in the introduction

ACADEMIC WRITING
25 Focussing on the issue (omitting personal pronouns)
26 Appropriate use of abbreviations
27 Avoiding contractions (e.g., don’t)
28 Avoiding extremeness (e.g., use of must)
29 Avoiding slang, jargon and clichés
30 Use of words with precise meaning
31 Use of objective language
32 Balanced use of passive forms

IDEA PRESENTATION
33 Appropriate use of markers (e.g., firstly)
34 Appropriate use of linking devices (e.g., however)
35 Flow of ideas
36 Paragraph unity
37 Overall unity
38 Paragraph coherence
39 Overall coherence
40 Appropriate length of paragraphs
41 Completeness of the sentences
42 Relevance of conclusions with the discussion
43 Drawing effective conclusions

MECHANICS
44 Paper format
45 Grammar
46 Spelling
47 Punctuation
48 Vocabulary selection
49 Use of tables and figures
50 Length of the paper

Evaluation criteria
Point(s)

Evaluation criteria
Point(s)

Evaluation criteria
Point(s)

Evaluation criteria
Point(s)