Scoring source-based writing: Challenges & Opportunities

Atta Gebril

The American University in Cairo
Outline

- Introduction
- Why source-based writing?
- Challenges
- Opportunities
- Research agenda
Introduction

- The writing assessment field has been intensively using **independent tasks (IT)** to assess the writing skills of university students.
- Independent tasks refer to tasks in which test takers are required to produce a text based on a certain prompt without using any sources.
- Many tests started using **integrated tasks** (e.g. TOEFL iBT, CAEL).
- **Reading-to-write** tasks refer to tasks in which students depend on information from a reading source to produce a text.
- **Listening to write** tasks refer to tasks in which students depend on information from a listening source to produce a text.
- **Source-based writing**
Why integrated tasks?

- Many **disadvantages** of IT
- Independent tasks are seen as a “**snapshot** approach” to writing assessment (Hamps-lyons & Kroll, 1995).
- Lack of **background knowledge** (Weigle, 2004)
- Fairness concerns (Plakans & Gebril, 2014)
- Independent tasks do not reflect the current **practices** in university classes (Cho, 2003)
Why reading-to-write tasks?

Research has shown that academic writing tasks are rarely done without using sources as a basis for writing (Weigle, 2002, 2004; Cumming, Kantor, Powers & Taylor, 2000; Leki & Carson, 1997; Hamp-Lyons & Kroll, 1996).
# University writing tasks

<table>
<thead>
<tr>
<th>Study</th>
<th>Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Elben (1983)</td>
<td>Analytical papers, abstracts, essays, lab reports, case reports, technical reports, and book reports</td>
</tr>
<tr>
<td>Bernhardt (1985)</td>
<td>Expository essay, critique, library research, argumentative essay, papers, book review, creative response, case study, lab reports.</td>
</tr>
<tr>
<td>Horowitz (1988)</td>
<td>Essay, summary of / reaction to a reading, case study, synthesis of multiple sources, research project.</td>
</tr>
</tbody>
</table>
Challenges
Demanding nature of integrated tasks

Integrated tasks are “more complex and more demanding than traditional stand-alone or independent tasks, in which test-takers draw on their own knowledge or ideas to respond to a question or prompt” (Brown, Iwashita, & McNamara, 2005, p. 1).
Demanding nature of integrated tasks

“Discussions usually arise among writing practitioners about rating-related issues, such as assessment criteria, discourse synthesis, and ownership. Critical in these debates are questions pertaining to how raters or instructors go about assigning scores to integrated tasks, what criteria they should use, and what constitutes appropriate textual borrowing practices.” (Gebril & Plakans, 2014)
Scoring Validity

Issues of interest in scoring validity investigation include:

- Assessment criteria used by raters
- Score reliability
- Scale accuracy
Criteria used by raters

- Raters may assign similar ratings for totally different reasons in spite of the fact that they are using the same scoring rubric (Douglas and Selinker, 1992).
- The relationship between the writing sample and the scoring rubric remains obscure (Lumley, 2002).
- Do raters use the same criteria to score different task types?
- Strong correlation between Integrated Tasks & Independent Tasks: Any interpretations?
Reliability concerns (Gebril, 2010)

Personal characteristics
Task
Test administration conditions
Rater
Rating scale
Writing mode / genre

Writing ability

Writing Score
Upshur & Turner (1999):
Score Reliability

• Some researcher voiced concerns with regard to score **reliability** (Lee & Kantor, 2005).

• Watanabe (2001); Lee & Kantor, 2005; Gebril, 2006) showed that source-based writing tasks could **yield as reliable scores** as independent tasks.
Score reliability of IT tasks (Gebril, 2009)
Score reliability of reading-to-write tasks (Gebril 2009)
Table 4
Source use concerns as reported by the two raters.

<table>
<thead>
<tr>
<th>Locating source use</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Checking for source use</td>
<td>“Does she incorporate? She doesn’t incorporate anything from…”</td>
</tr>
<tr>
<td></td>
<td>“The readings are used.”</td>
</tr>
<tr>
<td>Distinguishing between the writer’s language from the source materials</td>
<td>“So is she talking about the reading specifically or just her own… her own… knowledge of what she remembers?”</td>
</tr>
<tr>
<td>Linking exact ideas/references (names) to source texts</td>
<td>“Let me check the source.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Citation mechanics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct format of citation (misplaced quotation, wrong punctuation, etc.)</td>
<td>“OK, the author missed one quote mark, quotation mark.”</td>
</tr>
<tr>
<td></td>
<td>“but for this one it’s not… entire… it’s not entirely copies of the sentences.”</td>
</tr>
<tr>
<td></td>
<td>“There’s some citation issues — the person didn’t really cite anything— I mean the person uses the source information however didn’t cite, didn’t give a reference, didn’t give the author credit for it.”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Quality of source use</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuracy of source use</td>
<td>“… inaccurate important ideas.”</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Relevance of source use</td>
<td>“So I think it contains relevant information from the reading”</td>
</tr>
<tr>
<td></td>
<td>“… just giving some examples and… some quotes.”</td>
</tr>
<tr>
<td></td>
<td>“There is just not much use of the readings,”</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Clarity of source use</td>
<td>“… when the source is… when the part that is understandable is mainly from the source. But the part that is not so clear is confusing is from the author himself or herself.”</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Appropriateness textual borrowing strategies</td>
<td>“They’re just quoted and kind of… not really incorporated.”</td>
</tr>
<tr>
<td></td>
<td>“So unnn she nicely incorporates the quote from the readings.”</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Effectiveness of source use</td>
<td>“and that… these citations support the argument for global warming.”</td>
</tr>
<tr>
<td></td>
<td>“ok so good place to cite a source.”</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>Overuse of source materials</td>
<td>“The second paragraph is mainly quotes from the essay—quotes from the sources.”</td>
</tr>
</tbody>
</table>
What raters think?

• I’m not sure they were intentionally copying or just unaware of how to properly cite direct quotes. I detected it because I was familiar with the source texts. And, sometimes, all of a sudden a writer who struggled with grammar/structure produced a perfectly structured sentence. This was a clue also. It didn’t affect scoring in a significant way, since I just assumed they were quoting (not copying) and didn’t know how to do it correctly. I guess it affected how I score when I considered how well they incorporated the source texts—if they even knew how to cite appropriately.” (Gebril & Plakans, 2014).
Discourse synthesis issues.

- Difficulty in distinguishing between language cited from source materials and language produced by writers.

- Is it a citation problem? Plagiarism? Something else?

- This problem is less of an issue at lower levels since the interlanguage of writers at these levels could be easily distinguished from the cited statements.

- Difficulty in assigning a score to essays with high incidents of quotations, which is usually expected at lower levels as suggested in the literature (Gebril & Plakans, 2009, 2013).
Opportunities

- Increasing interest in content-related issues
- Adding weight to issues of writing beyond the fundamental language structure by shifting focus to discourse development
- More interest in source integration quality
  - While the attention paid to source use at lower levels focused on the presence/lack of source materials and citation mechanics, at higher levels the attention was shifted toward other issues, such as the adequacy of source use, clarity of citation information, and effectiveness of source use. Although some of these areas were reflected in the scoring rubric, current scoring rubrics need to be refined to better reflect these source integration issues.
Lower level courses might best focus on the basics of sources use such as the how and why of citation as well as reading skills such as comprehension and scanning.

At mid range levels, more complex issues such as selecting and paraphrasing ideas from sources could be taught.

At advanced levels, instruction can tackle the trickier aspects of source use: relevance of source ideas to author’s argument, including enough source support (and not too much), and integrating while maintaining a smooth flow in writing.

The writer’s voice is an important issue that should be reflected in scoring rubrics.
Appropriateness of textual borrowing

- It is important for raters and writing instructors to have a sense of what is considered appropriate textual borrowing practices.
- Be aware of the instructional and cultural contexts in which L2 writers operate. We need to understand that the concept of plagiarism is perceived differently in various cultures as argued by Pennycook (1996).
- This understanding requires not only educating raters and writing instructors, but also students and other stakeholders.
- Positive washback in writing classes.
Rater training & scoring rubrics

- Rater training sessions should go beyond mechanical issues to address source use accuracy, effectiveness, clarity, and overall integration quality.

- These issues should also be reflected in integrated scoring rubrics with clear guidelines about how to address them at different proficiency levels.

- We argue that scoring rubrics for integrated tasks should have progressively different descriptors for source use, with lower levels addressing mechanics and higher levels addressing quality of source use.
Future research?

- Criteria raters use when they score different types of integrated tasks
- Development and validation of new integrated writing scoring rubrics
- Discourse synthesis features across different proficiency levels
- Washback in Writing classes
Thanks for your attention!