A closer look at integrated writing tasks: Towards a more focussed definition for assessment purposes

Ute Knoch\textsuperscript{a,∗}, Woranon Sitajalabhorn\textsuperscript{b,1}

\textsuperscript{a} Language Testing Research Centre, School of Languages and Linguistics, University of Melbourne, Parkville 3010, Victoria, Australia
\textsuperscript{b} School of Languages and Linguistics, University of Melbourne, Parkville 3010, Victoria, Australia

\textbf{A R T I C L E   I N F O}

\textbf{Article history:}
Received 18 March 2013
Received in revised form 4 September 2013
Accepted 4 September 2013
Available online 6 October 2013

\textbf{Keywords:}
Integrated tasks
Integrated writing tasks
Writing assessment

\textbf{A B S T R A C T}

The use of integrated tasks is increasingly common in second language assessment in both high stakes and classroom contexts. However, because of the vague definition of what constitutes an integrated task, the types of tasks included in this area and the assessment of the resulting writing is highly varied. In this paper, we argue for a better definition of the term ‘integrated writing task’. We start by reviewing current definitions available in the literature, discuss the construct underlying integrated tasks and then propose a number of features which we feel need to be present in integrated tasks. We then propose a new, narrower definition of the term. We hope that this will result in better task design, improved assessment outcomes for test takers, more detailed test specifications, better replicability of research and a shared research agenda.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Integrated writing tasks have in recent years grown in popularity and have been adopted widely as tools to measure second language learners’ writing ability. The integrated task of the writing section of the Internet-based Test of English as a Foreign Language (TOEFL iBT) is one example. According to

\textsuperscript{∗} Corresponding author. Tel.: +61 3 83445206.
\textit{E-mail addresses:} uknoch@unimelb.edu.au (U. Knoch), w.sitajalabhorn@student.unimelb.edu.au (W. Sitajalabhorn).
\textsuperscript{1} Tel.: +61 423812155.

1075-2935/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.asw.2013.09.003
Cumming et al. (2006), the Educational Testing Service introduced the integrated task on the TOEFL to “improve the measurement of test takers’ writing abilities, create positive washback on teaching and learning as well as require test-takers to write in ways that are more authentic to academic study” (p. 1). Other authors have argued that such tasks address authenticity and validity issues, improve test fairness, and provide learners and test takers with content and language support for their writing (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Feak & Dobson, 1996; Fox, 2004; Leki & Carson, 1997; Plakans & Gebril, 2012; Raiines, 1998; Read, 1990; Weigle, 2004). However, Plakans (2012) cautions that “while benefits exist for integrating skills in assessment, numerous challenges surface that warrant attention when developing and using such test tasks” (p. 250).

One of the fundamental challenges, we believe, lies in the lack of a clear and sufficiently thorough definition of integrated tasks. Whilst several studies (e.g., Ascención Delaney, 2008; Esmaeili, 2002; Gebril, 2010; Plakans, 2008, 2009a, 2009b, 2010; Plakans & Gebril, 2012; Yang, 2009, 2012) have been conducted to gain more insight into different aspects of integrated writing tasks, to date the tasks themselves have not been clearly defined (see also Cumming, 2013; Yu, 2013). A more focussed, slightly narrower definition of these tasks could improve test design and improve replicability of research including a coherent research agenda.

This paper will start by exploring some current definitions of integrated writing tasks, discussing what we feel are clear shortcomings. We will then describe some of the myriad of task types currently used under the umbrella of integrated writing tasks. Following a discussion of the construct underlying such tasks, we will describe features of tasks and rating scales which need to be present for a task to be classified as ‘integrated’. We conclude the paper by providing our own definition of integrated tasks and a framework for validation of such tasks.

2. Existing definitions of integrated writing tasks

Although quite a few studies on integrated writing tasks exist, only a small number of them touch upon the issue of task definition. For example, Ascención Delaney (2008), examining the construct of reading-to-write tasks, explains that they are “instructional tasks that combine reading and writing for various educational purposes” (p. 140). Plakans (2009b) has expanded the definition a little further by including other skills. To her, integrated writing tasks are those which “elicit writing performance that involve other abilities such as reading or listening” (p. 252). In her other work, she has provided a very broad definition of integrative language tasks, describing them as “tasks that require more than one skill for completion” (Plakans, 2012, p. 249). In the most detailed definition we found, Cumming et al. (2005) describe integrated tasks as those in which test takers are required to “produce written compositions that display appropriate and meaningful uses of and orientations to source evidence, both conceptually (in terms of apprehending, synthesising, and presenting source ideas) and textually (in terms of stylistic conventions for presenting, citing, and acknowledging sources)” (p. 34).

3. Integrated writing tasks currently in use

As is evident from the definitions provided above, integrated writing tasks have been defined broadly, and the definitions appear to centre on the issues of skill amalgamation and source use. With such broad definitions, it is not surprising that the term integrated task has been applied to a copious array of task types. In this section, we will present some of the writing tasks currently described in the literature as integrated tasks. It will be seen that these vary considerably in their features.

The most common task type found in the literature on integrated writing tasks is one that provides reading material and requires learners to compose a text based on the information present in the material, thus known as a reading-to-write task. Several different versions of reading-to-write tasks can be found in the literature. Firstly, summary tasks have been used in studies by Ascención Delaney (2008) and Yu (2007, 2008, 2009, 2010). These require test takers to write a summary of a passage.

---

2 We will only focus on assessment tasks (and not all pedagogical tasks) in this paper, although the principles may also apply to those.
A response essay, employed as a research instrument in Ascención Delaney’s (2008) study, also falls into this task type (Grabe & Zhang, 2013; Plakans, 2012). To complete this task, learners/test takers summarise a reading passage provided and then express their views on the topic or respond to predetermined questions. A variant of this task type involves a reading stimulus with two passages. Test takers might be asked to write a synthesis of several source texts (e.g. Grabe & Zhang, 2013) or present two opposing views on the same topic and then consider their position/stance and write an essay to defend it, combining the main ideas from the passages with their own opinions and examples.

Whilst the tasks above rely solely upon reading input, other integrated tasks also incorporate listening input. As described by Alderson (2009) in a test review, one such example is the TOEFL iBT integrated writing task. Test takers read a passage and listen to a lecture on the same topic (note taking is optional); afterwards, they summarise the main points made by the lecturer and state how the points relate to the content of the reading passage.

In an attempt to establish the writing framework for the TOEFL iBT, Jamieson, Eignor, Grabe, and Kunnan (2008) also proposed a situation-based integrated writing task. The task would have test takers read or listen to a brief exchange between or among people and “describe a problem and propose a solution or synthesise and summarise information in writing”, and their written response must be “appropriate in terms of the content, setting, and purpose” (p. 79). Although this task did not make its way to the TOEFL iBT, a similar task is employed by the University of Cambridge ESOL Examinations (Cambridge ESOL) in their writing section of the Certificate in Advanced English (Cambridge ESOL, 2012).

All the tasks described above require learners to transform the language of the input material to accomplish the writing task. This means that the language of the input material is crucial to the accomplishment of the task. However, Plakans (2012) notes the existence of another type of integrated task in which “input texts are provided for test takers but not required in the response” (p. 256). An example, given by Plakans (2012), is a task in which learners/test takers listen to a conversation with two people discussing different genres of movies and trying to make a decision as to which movie they should watch. The task instructions then ask them to suggest a movie they like and explain why they make that suggestion. In such a task, even though the input material is important for task completion, the language of the input is not needed to complete the task and the topic and the written output is generated mostly by learners/test takers themselves.

Finally, for Yang (2012), a graph-writing task is also regarded as an integrated writing task. Instead of reading and/or listening materials, participants in Yang’s (2012) research were provided with a line graph and a pie chart and asked to write an essay evaluating the data and explaining their relationships. Similar tasks are employed by IELTS (Task 1) and the Diagnostic English Language Needs Assessment (see e.g. Elder & von Randow, 2008; Knoch, 2007; Read, 2008).

As illustrated in the above review, a wide variety of task formats fall under the umbrella of integrated tasks. Some require only one receptive skill (with input from one or multiple sources), whilst others call for test takers to read and listen before writing takes place. Some provide text-based input; others auditory; still others visual or the combination of these. Some demand the presence of information or content from source materials in a written product, yet others do not.

4. Test construct

Considering the many test task types currently used under the umbrella of integrated tasks, it is important to give some consideration to the construct underlying integrated writing tasks. This is vital, as the quality of the interpretations based on the test scores are dependent on the initial construct definition. A clear construct definition plays a pivotal part in an argument-based approach to validation (see for example Chapelle, 2012; Chapelle, Enright, & Jamieson, 2008; Chapelle, Enright, & Jamieson, 2010; Kane, 1992, 2006, 2012; Messick, 1989; Xi, 2008).

Unfortunately, a complete theory or model of writing from sources is not available at this current point in time (see e.g. Ascención Delaney, 2008; Hirvela, 2004), and research into how learners, especially those learning a second language, acquire the ability to write from sources is in its infancy. Most work in this area has been undertaken with adult learners, usually in university or pre-university contexts. Currently, then, we must look elsewhere for guidance on this construct.
Scholars working on a better theoretical understanding of the demands of integrated writing tasks have studied test takers’ cognitive processing. This line of research has revealed that test takers use reading materials to mine and select ideas, shape and support their opinions, seek language support, and help organise their writing (Leki & Carson, 1997; Plakans & Grbrl, 2012). Grabe (2003, pp. 244–245), for example, indicates that, to complete these source dependent writing tasks, writers are required to make a number of decisions, including how much and which information should be selected from the text, how the information will fit task demands and writer goals, how accurately the information from the source texts should be represented and what formal mechanisms should be used for transforming the textual information to fit the final written product.

In a study representative of those undertaken in the second language context, Plakans (2009a), following the discourse synthesis framework developed by Spivey and King (1989), has investigated discourse synthesis sub-processes employed by writers during the production of reading-writing tasks written for a university English placement test. Based on Spivey and King’s (1989, p. 11) definition, a discourse synthesis entails three essential sub-processes: organising (when writers think about the overall structure of their writing and the structure of the readings), selecting3 (when readers read and choose ideas from the readings), and connecting (this includes linking ideas in the writing and connecting ideas in the readings with the writer’s own ideas’). The results revealed that participants with higher English language proficiency (judged by their TOEFL scores) used the sub-processes of organising, selecting, and connecting considerably more frequently during their composing processes than their less proficient cohorts did. According to Plakans (2009a), participants whose overall TOEFL scores were lower paid more attention to language difficulties and writing, resulting in poorer performance of discourse synthesis. Based on these results, Plakans (2009a) suggests the inclusion of the organising, selecting, and connecting abilities as a construct of academic writing, maintaining that this would enable us to interpret the scores on integrated writing tasks more validly.

In sum, research has shown that the construct underlying test takers’ behaviour when completing integrated writing tasks includes these prominent elements:

1. Mining/selecting the input text(s) for ideas to be used.
2. Synthesising ideas from various sources or summarising from one source.
3. Transforming the language used in the source text(s).
4. Choosing the organisational structure to be used in the writing (which is often different from the structure of the input text).
5. Connecting the ideas in the writing; connecting ideas in the reading with their own ideas.

Apart from the work done to understand the cognitive processes writers engage in when completing integrated writing tasks, research has also investigated the factors that affect performance on such tasks. Studies have examined the influence of different test taker variables on the performance on integrated tasks (see for example Enright, Bridgeman, & Cline, 2002; Lewkowicz, 1994; Risemberg, 1996; Trites & McGroarty, 2005; Watanabe, 2001) and have shown that both reading and writing ability by themselves contribute to the total variance in integrated writing tasks, with writing ability being the stronger predictor.

Other research has focussed task variables that influence task difficulty (e.g. Ascención Delaney, 2008; Cho, Raijen, & Novak, 2013; Ruiz-Funes, 1999; Spivey, 1997). This research is still in its infancy but it has been shown that certain aspects of the prompt/task characteristics (e.g. the nature of the topic, the length and complexity of the source material, the mode of presentation of the input material and the type of task required of the writer) have an influence on the writing performance.

The brief discussion of the construct of integrated writing task above4 shows that more work in this area is needed. Most research thus far has focussed on reading-to-write tasks and on learners in university settings. Many aspects of test taker cognitive processing on other types of integrated tasks,

3 Please note that this sub-process has also been referred to as ‘mining’ (see Esmaeili, 2002; Greene, 1992; Hirvela, 2004; Plakans, 2009b).
4 Please refer e.g. to (Ascención Delaney, 2008) for a fuller discussion of some of the issues.
the similarity or differences of such processes on assessment tasks and real life tasks, and differences in discourse across tasks with a variety of test taker populations and in different contexts are yet to be investigated.

5. Towards an improved definition of integrated writing tasks

To advance our understanding of the construct underlying integrated writing tasks, we argue that a more thorough definition of these types of tasks is needed. Before we offer such a definition, we will now turn to two areas of assessment which are directly related: (1) task design and (2) rating scale design. Both these areas are inextricably related and therefore task and rating scale design need to be an iterative process when integrated writing tasks are conceptualized and designed. Unfortunately, we believe that some tasks currently used under the umbrella of integrated tasks are not truly integrated according to our more closely-specified definition. For this reason, we will discuss a number of features of writing tasks which make them suitable to be viewed as integrated tasks under the slightly narrower definition we will propose later in this paper. We will first discuss task features and then rating scale features.

5.1. Task features

The review of the myriad of tasks currently used under the term integrated writing tasks has shown that the tasks vary significantly. However, to make advances in our understanding of the construct underlying integrated tasks and work towards a fuller theory of writing from sources, we argue that integrated writing tasks need to display two features: (a) the input material needs to include a significant proportion of language and, directly following from this, (b) the task needs to require that the language in the source material is used and transformed to complete the writing task.

We would like to argue that a writing task cannot be defined as an integrated task unless the input material contains a significant proportion of language. Therefore, the graph writing task described by Yang (2012) as an integrated task does not meet this criterion and cannot be considered an integrated task. If the stimulus is not language rich, it is likely to be serving merely as a source of ideas or content that may trigger the output but will not involve any meaningful integration of skills. What constitutes a significant proportion of language is dependent on the type of test as well as the test taker population or student group the task is aimed at and the context in which it is used. The input material would be much shorter in the case of a test for young learners with limited literacy or a test targeted at beginning learners of a language than a test designed to measure the writing ability of students in an academic setting. However, the stimulus materials need to provide sufficient language (either in written or audio format) to allow writers to produce sufficient text to be rated by assessors.

The second task feature we would like to advocate is that test takers need to be required to transform the language from the input material in some way for their written output. Depending on the task, this can require writers to summarise the language by selecting the most important ideas and using paraphrasing techniques. It can also mean that the writer is required to select different points of views to critique or agree with and elaborate on. Therefore, the language in the source material cannot solely be used to ‘set the scene’ for a task scenario in which learners completely rely on their own language resources and ideas as is the case with the task described by Plakans (2012).

5.2. Rating scale features

While the issues relating to task design discussed above are clearly essential, it is equally important that the linguistic demands of these tasks are adequately reflected in the rating scale used to rate the writing product as rating scales have the function of representing the test construct. It is therefore important that the writing rubrics used ‘measure’ the construct of skill integration and provide a

---

5 Please note that although we argue that graph writing tasks are not integrated tasks, we acknowledge that these tasks are part of the construct of academic writing and therefore an important, valid skill to assess.
working definition for the users of the scale. If a rating scale does not adequately account for these aspects, even a well-designed task which resembles an integrated task in other respects will underrepresent the integrated writing task construct.

Based on the studies reviewed in the discussion of the construct of integrated tasks above, a rating scale should evaluate test takers’ ability to mine, select, organise and connect ideas in the written product. In fact, the processes outlined above are essential when accounting for skill integration. The rating scale should also account for the transformation that has taken place in the language from source text to the final written product. This is an important point as unless the language is transformed, language use cannot be evaluated by a rater. For example, in a rating scale designed to rate summaries, this would be reflected in the evaluation of the effectiveness of paraphrases (Plakans, 2009b). For integrated writing tasks developed for the academic domain, how appropriately writers acknowledge the sources might be another aspect that can be incorporated into a rating scale.

When developing a scale, test developers need to choose whether to use a holistic rating scale or an analytic rating scale. Regardless of the scale type chosen, the criteria should reflect what test takers are required to do with the source material (e.g., selecting, transforming, organising, acknowledging source texts). A good example of a holistic scale for integrated tasks are the TOEFL integrated writing rubrics. In this task, test takers are required to integrate information from both a reading and a listening text into their written product. The Level 3 descriptor can be seen below.

**Level 3:**

A response at this level contains some important information from the lecturer and conveys some relevant connection to the reading, but it is marked by one or more of the following:

- Although the overall response is definitely oriented to the task, it conveys only vague, global, unclear, or somewhat imprecise connection of the points made in the lecture to the points made in the reading.
- The response may omit one major key point made in the lecture.
- Some key points made in the lecture or the reading, or connections between the two, may be incomplete, inaccurate, or imprecise.
- Errors of usage and/or grammar may be more frequent or may result in noticeable vague expressions or obscured meanings in conveying ideas and connections.


The TOEFL writing rubrics for the integrated task make reference to most of the aspects discussed above. The Level 3 descriptors make no direct mention of language transformation, but this issue is taken up at lower levels in the rubrics when referring to ‘... merely copies sentences ...’. Clearly, transformation of the language, for example by paraphrasing, is necessary in a language assessment.

If an analytic scale is chosen, features of integrated writing can be represented in several categories. For example, a category on content can reflect whether the ideas from the source text(s) were selected appropriately. A category on organisation/coherence can comment on the organisation of the ideas from the source text(s) in the test takers’ essays.

Table 1 outlines possible example descriptors (depending on task requirements) for high and low achieving writers.

Finally, as suggested by Fulcher (1996, 2003), the descriptors of each score level should be designed and determined based on samples of actual student writing so that the scale is a true reflection of student performance at different writing levels of proficiency (see e.g. Inoue, 2009; Ohkubo, 2009).

6. Towards an improved definition of integrated writing tasks

This paper has attempted to review the existing definitions of integrated tasks and argues that a carefully specified definition has yet to be established. What we have currently as definitions of
Table 1
Example descriptors for analytic rating scale.

<table>
<thead>
<tr>
<th>Content</th>
<th>Language</th>
<th>Organisational structure</th>
<th>Cohesion</th>
<th>Acknowledgement of sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Major ideas from the input materials have been selected</td>
<td>The language of the stimulus material has been appropriately transformed using paraphrasing techniques</td>
<td>The structure of the written text is appropriate to the task requirements</td>
<td>The ideas from the input have been successfully integrated with each other and the ideas from the writer</td>
<td>The author appropriately cites the source texts</td>
</tr>
<tr>
<td>Few ideas from the content have been included</td>
<td>Large stretches of language have been copied from the input material</td>
<td>The text structure of the written product does not mirror task requirements</td>
<td>The ideas taken from the input seem disjointed and have not been integrated well with the writer’s own ideas</td>
<td>Source texts have not been cited appropriately</td>
</tr>
</tbody>
</table>

integrated tasks are relatively vague and often so broad that virtually all the tasks with any type of stimulus can be regarded as integrated. In this paper, we have briefly examined the construct of integrated writing tasks and then described features of writing tasks which we consider essential for tasks to be counted as integrated. We have argued that for a writing task to be integrated, (1) the source material should include a significant amount of language (which therefore excludes input based on visual stimuli only), (2) the writing product needs to draw on the ideas presented in the input text(s), and (3) the language presented in the source texts needs to undergo some language transformation before being used in the written product. Furthermore, the rating scale used to assess the written products needs to reflect these aspects to fully capture the construct being assessed by the task. The rubrics should provide raters with descriptors which are reflective of what test takers are required to do with the source material and should ideally be developed based on a review of sample performances by test takers.

Based on the summary provided above, we would like to propose the following expanded definition of integrated writing task:

Integrated writing tasks are tasks in which test takers are presented with one or more language-rich source texts and are required to produce written compositions that require (1) mining the source texts for ideas, (2) selecting ideas, (3) synthesising ideas from one or more source texts, (4) transforming the language used in the input, (5) organizing ideas and (6) using stylistic conventions such as connecting ideas and acknowledging sources. The rating scale used to grade such compositions needs to take account of these features specific to integrated writing tasks.

7. Conclusion

We hope that this new definition addresses the lack of a ‘coherent manifesto for the design of integrated writing tasks’ (Cumming, 2013) and therefore will improve future writing task and test specification design and promote a consistent research agenda (see also Yu, 2013) so that findings can be compared across different assessment contexts and bridge some of the research gaps still existing in this area. For this same reason, we have chosen to provide a focused definition of integrated tasks, rather than a typology of different task types. We argue that this approach will be more helpful to further a coherent research agenda.

For test validation activities, we propose an argument-based approach to validation (Chapelle, 2012; Chapelle et al., 2008, 2010; Kane, 1992, 2006, 2012; Messick, 1989; Xi, 2008) with its series of inferences which help test developers to systematically collect evidence on a variety of aspects. Validation research to support the evaluation inference would, for example, collect evidence to establish whether the rating scale does reflect the relevant knowledge and skills displayed by learners completing the integrated task. Rater verbal protocols as well as an analysis of test taker discourse (see for example Cumming et al., 2006; Plakans & Gebril, 2013) are commonly used methods for this purpose. A quantitative analysis of test scores, using, for example, multi-faceted Rasch measurement,
could show how reliably raters are able to apply the scoring rubric and whether any of the criteria on the rating scale commonly result in rater bias. If an automated scoring system is used to score the performances on an integrated writing task, then evidence needs to be collected to show that there is no construct-underrepresentation in the nature the essays are evaluated. Comparisons of the scoring accuracy of the automated system and human raters would also be important.

Evidence for the generalization inference will focus on collecting evidence to show equivalence of multiple ‘equivalent’ versions of the same prompt. This is not a simple undertaking as often the entire writing form consists of only one or two prompts. Equivalence is usually established through rigorous item development and review processes although larger-scale tests can establish effects of prompts difficulty using quantitative means (see e.g. Cho et al., 2013).

To collect evidence for the explanation inference, researchers would collect data which shows that writers engage in processes which are representative of those used on integrated writing tasks in the target language use domain. This would usually be done by collecting verbal protocols of test takers engaging in such tasks as well as a comparison of the discourse produced in response to the integrated task in comparison with that produced on similar real-world tasks. Questionnaires could also be administered to test takers to gain a deeper understanding of the processes and strategies used by test takers during task completion.

Evidence to support the extrapolation inference could include two types of evidence: judgments that the writing tasks are representative samples of the target language use domain, and evidence that the test scores on the integrated writing task are highly correlated with scores on the criterion measures.

References


**Ute Knoch** is the Director of the Language Testing Research Centre at the University of Melbourne. Her research interests are in the areas of written assessment, assessing languages for specific purposes, diagnostic assessment, placement testing and rating scale design and validation.

**Woranon Sitajalabhorn** is a PhD candidate at the School of Languages and Linguistics, the University of Melbourne. He is currently conducting a study on the assessment of summary writing. His research interests lie in the areas of written assessment and the development of writing rating scales.