A Comparison of Writing Performance on Independent and Integrated Writing Tasks

Rahil Sheibani, Ph.D. Candidate, Qeshm Branch, Islamic Azad University, Qeshm, Iran
rahil.sheibani@yahoo.com
Alireza Ahmadi*, Associate Professor, TEFL, English Department, Shiraz University, Shiraz, Iran
arahmadi@shirazu.ac.ir

Abstract
Researchers and scholars have been attracted by the idea of using integrated writing task along with independent writing task to best assess the EFL learners' writing competence. This study was conducted to compare the writing performance of EFL students in integrated and independent writing tasks. It also aimed to find out if writing performance varies with task types. A number of thirty Iranian EFL students participated in this study, and each student wrote on two writing tasks of IELTS Academic module. The written essays were rated by three experienced raters using IELTS rubrics. Paired sample t-tests revealed that test takers performances did not differ significantly across these two tasks. Besides, two one-way ANOVAs indicated the difference between the raters in rating integrated writing task while they were not performed differently in rating independent task. The results are followed by a number of suggestions with the aim of improving the EFL students' performance in writing skill.

Keywords: Writing competence, academic writing tasks, Iranian EFL learners, independent writing task, integrated writing task

Introduction

Speaking and writing, the two productive skills in second language, are crucial abilities for communicative reasons (weigle, 2009). Writing is the most difficult and challenging language skill because it requires conscious effort and practice in creating, developing, and analyzing ideas (Emig, 1997). Therefore, L2 language learners usually encounter difficulties in developing their writing (Evans et al., 2010).

In educational context, students' achievement, learning and intelligence are judged according to their writing performance (Ghalib & Al-Hattami, 2015). As Emig (1997) suggested, learning writing requires only formal and systematic training. Also, writing instructors have noticed that for many students it is a difficult skill to acquire even in their native language. ESL/EFL research has shown that direct assessment is complex and challenging (Barkaoui, 2008; Huang, 2011; Huang & Foote, 2010; Huang & Han, 2013; Sakyi, 2000). Besides, teachers argued that as writing is one of the most important subjects taught in schools, it has to be accompanied with effective assessment (Coffman, 1971). Hughes (2003) stated that the best way to assess the writing ability is to get the students to write, accordingly writing assessment includes one or more writing tasks or instructions that tells test takers what to write, and a means of educating these writing samples.

Indirect testing can help instructors to test individual aspects of language related to writing. Also, it can easily measure points of student error, so test evaluators can give feedback on students' problem areas. Criticism has been laid against indirect testing of writing proficiency since it just concerns about students' knowing about writing which does not able them to write effectively. In consequence, the indirect method of testing has no longer been considered an

accurate measure of writing ability (Brown & Hudson, 2002; Hughes, 2003). On the other hand, in direct composition tests students produce actual writing. Direct testing involves a prompt which places the students in a particular situation where they are required to use writing language to accomplish specific real communication objectives. Direct assessment is more inductive and requires evaluators to infer students' knowledge and understanding based on their ability to produce a meaningful piece of writing. However direct testing is more reflective of real writing ability, its means of evaluation are more subjective (Brown et al., 2002; Hughes, 2003). Educational community agreed that writing is a language skill that is best tested directly, rather than indirectly (Hughes 2003).

By the introduction of the direct writing tasks (Hamp-Lyons, 1991) and the emergence of the standardized tests including TOEFL and IELTS for admission or placement decisions in universities, researchers focused on assessing the academic writing abilities of university students (Cho, Rijmen, & Novak, 2013; Read & Hayes, 2003).The timed-essay independent task has been frequently used as the main method to test the university students' writing skills (Gebril, 2009). But many researchers criticized it because of its inadequacy in employing the academic writing construct (Cho, 2003; Gebril, 2006; Gebril & Plakans, 2009; Plakans, 2007; Weigle, 2002, 2004).

Conventional essay tests include writing tasks which lack theoretical validity and applicability in real-life contexts (Cho, 2003). Besides, independent tasks have problem in topic familiarity and the students' performance is affected by the assigned topic if they do not have sufficient background knowledge. Moreover, since independent writing tasks are decontextualized, the test takers cannot use any sources in producing their texts (Hamp-Lyons & Kroll, 1996). According to Gebril (2006), writing tasks require topic familiarity; therefore, lack of topic familiarity leads to underestimating test takers' writing competence (Hamp-Lyons & Kroll, 1996).

To overcome the shortcomings of independent tasks, integrated tasks have been recommended as an alternative, or addition, to independent tasks (Gebril, 2009). In independent writing tasks, the test takers have to depend on their own knowledge of topic to write a text, while in integrated tasks, they can benefit from provided sources. Academic writing tasks require using reference sources as a basis for writing (Gebril, 2009; Hamp-Lyons & Kroll, 1996; Leki & Carson, 1997; Weigle, 2002, 2004).

The most accepted definition of integrated writing tasks was proposed by Weigle and Parker (2012), and Plakans and Gebril (2013). They defined it as writing tasks in which the writer will respond by combining the concepts presented in several source texts. The definition of integrated writing tasks has been redefined by Knoch and Sitajalabhorn (2013) as a process requiring the test takers to scrutinize the prompts in looking for ideas, combine the different concepts, organize their ideas, and using conventional writing knowledge to write the final draft.

It is claimed that combining writing skill with other language skills is considered beneficial. Inclusion of written or spoken source texts in integrated tasks provides sufficient content, authenticates and validates the measures of testing academic writing ability (Plakans, 2008). Besides, Plakans and Gebril (2012) maintained that combining writing skill with other language skills such as reading and speaking can create a view that language is holistic rather than componential. Leki and Carson (1994) were convinced that EAP writing classes require to pass away from stand-alone writing tasks in which students tap into their own opinions and experiences and go toward tasks that encourage learners to combine their own opinions and experiences with external sources of information and argument.

Plakans (2007) pointed out that in integrated tasks, reading a source text before writing can provide a common platform for test takers who bring different cultural and individual

characteristics to the exam room. Additionally, it is claimed that integrated tasks reflect authentic academic tasks (Wiegle, 2004), and improve the strategic competence since completing such tasks requires applying different strategies (Plakans, 2009).

Integrated writing tasks have been employed frequently for assessing academic writing ability in standard proficiency assessments. TOEFL and IELTS tests have considered different types of writing tasks for the writing section of the new tests including integrated writing task and independent writing task. In integrated tasks, test takers are asked to read and/or listen to one or more texts (e.g., a reading passage, a lecture) and then create a written response, whereas in independent writing tasks they are required to write about a topic based on their personal experience and/or general knowledge without referring to any other sources (Cumming et al., 2000; Jamieson, Eignor, Grabe & Kunnan, 2008). Independent tasks are assumed to provide a measure of writing as an independent skill while integrated writing tasks entail the use of two or more language skills at the same time (Jamieson et al., 2008; Read, 1990). Integrated writing tasks require writers to synthesize information from other sources in their writing (Gebril & Plakans, 2009, 2013). Researchers declare that integrated tasks like the actual practices in academic contexts require discourse synthesis which is a common exercise in university writing (Gebril, 2009; Horowitz, 1986; Moore & Morton, 1999; Plakans, 2008, 2009). These researchers argue that, the university students are required to write considerable amount of source-based writing in a variety of genres, including research papers, reports, reaction papers, and case studies. Integrated tasks are more complex and more demanding than independent tasks and can enhance authenticity and better elicit the academic writing construct (Brown, Iwashita, & McNamara, 2005).

Review of Literature

Background of EFL/ESL writing points out that academic writing ability consists of independent or writing-only tasks and integrated or source-dependent tasks. Source-dependent tasks were firstly suggested by the Educational Testing Service (ETS) to the administrators of the Test of English as a Foreign Language with the aim of measuring writing ability more accurately (Plakans & Gebril, 2013; Weigle & Parker, 2012).

Besides, Plakans (2008) outlined the issues of rating, construct validity and authenticity in second language writing assessment as the pitfalls of writing-only tasks. Guo et al. (2013) scolded the independent writing task as a decontextualized writing activity. Review of writing instruction literature reveals that independent writing tasks have been employed in two major manifestations including feedback (error correction) and collaborative writing (Gholami & Alinasab, 2016).

Although several studies have examined the effects of task characteristics variation on L2 learners' writing performance (e.g., Clachar, 1999; Cumming, 1989; Krapels, 1990; Raimes, 1987), there is a dearth of study in writers' performance on independent and integrated writing tasks. As mentioned above, in response to independent writing tasks, test takers write about a topic based on their personal experience and/or general knowledge without assistance of any other sources, while for accomplishing an integrated tasks, test takers read and/or listen to one or more texts and compose a written response (Cumming et al., 2000; Jamieson et al., 2008). Integrated writing tasks make use of two or more language skills, while writing in independent tasks is considered as an independent language skill (Jamieson et al., 2008; Read, 1990). Inspecting writing processes of different tasks can reveal whether these tasks tap the same construct (Barkaoui, Brooks, Swain, & Lapkin, 2013; Lee & Kantor, 2005).

Some studies have inspected the effects of different writing tasks on different issues such as test scores (Lee & Kantor, 2005), writing processes (Plakans, 2008), and text features (Cumming et al., 2005). For example, Lee and Kantor (2005) by comparing writing tasks found high correlations among scores on independent and listening and reading-based integrated tasks; they found that, these tasks may be measuring the same underlying construct. In another study, Cumming et al. (2005) reported the significant differences across integrated and independent writing tasks in terms of various linguistic and discourse features. Guo et al. (2013) explored whether linguistic features can determine second language writing proficiency in integrated and independent writing tasks. They studied linguistic features such as lexical sophistication, syntactic complexity, cohesion, and basic text information in integrated and independent writing samples. The results revealed that, linguistic features can predict scores of both integrated and independent writing tasks. They concluded that evaluation of these writing tasks relies on similar and distinct features.

Some studies have inspected the writing processes of L2 learners when responding to integrated tasks (e.g., Cohen, 1994; Esmaeili, 2002; Plakans, 2009) or independent tasks (e.g., Clachar, 1999; Raimes, 1987), while Plakans (2008) compared L2 writing processes for both task types. She compared test takers' processes in composing two writing tasks. Differences were found across tasks and writers; the integrated task was more interactive while independent tasks required more initial and less online planning. Additionally, experienced writers were interested in writing with more interactive process. The results also specified that the respondents of independent task engaged in initial planning, while in integrated tasks they engaged in more online or during writing planning.

Integrated essay writing tasks have mostly been studied in the field of EFL/ESL writing assessment rather than writing instruction. Different effective factors like source text borrowing, scoring, reading-to-write construct, paraphrasing, strategy use and processes involved during writing were investigated (Plakans & Gebril, 2012; Hirvella & Du, 2013; McCulloch, 2013; Plakans, Weigle & Parker, 2012).

As IELTS tests included both independent and integrated writing tasks in its writing section, some studies focused on the writing section of IELTS. For instance, Golder, Reeder, and Fleming (2011) determined the appropriate band scores for admission into programs while Moore and Morton (2005) compared IELTS writing tasks with the university writing and rater variation in scoring was investigated by Lee and Kantor (2005). Other researchers worked on examining IELTS as an indicator of written proficiency levels (Ellis, Chong, & Choy, 2013), elaborating on IELTS gain scores (Elder, & O'Loughlin, 2003; Read & Hayes, 2003), and studying the washback effect of IELTS and the impact of preparation programs on candidates' performance (Green, 2007).

In a comparative study Moore and Morton (2005) compared the IELTS writing Task 2 (independent task) with a corpus of 155 assignment tasks collected at two Australian universities. The findings revealed that in spite of some similarities between this independent writing task, and the predominant genre of university study (the essay), some important differences also existed between these two as well. The results indicated that the type of writing that IELTS Task 2 elicits has more in common with nonacademic genres and as such cannot be considered appropriate in eliciting a sample of university writing.

In a quantitative study, Ellis et al. (2013) examined the communicative proficiency of student teachers. The writing proficiency levels were measured by the participants' IELTS writing scores. The findings showed that, the lowest scores were related to the writing section.

Furthermore, they had better scores in argumentative writing, while they had more problems in describing visual prompts like a table or a graph.

In another study, Cumming et al. (2001) investigated raters' decision-making behaviors while scoring independent and integrated writing tasks. The required data was elicited by conducting think-aloud protocols during scoring writing samples. The results showed that raters paid more attention to rhetoric and content in integrated tasks when compared to independent tasks, and they paid less attention to language. The authors attributed this increased attention to rhetoric and content to the nature of integrated writing, which requires more focus on source materials and students' use of them in their writing. The results also pointed out that the raters requested more guidance regarding the rating criteria while working on integrated tasks.

In another related project, Gebril and Plakans (2014) investigated how raters approach reading-to-write tasks, how they react to source use, the challenges they face, and the influencing factors on their scoring decisions. To yield these aims, they employed an inductive analysis of interviews and think-aloud data obtained from two raters. The results showed that raters focused more on judgment strategies rather than interpretation behaviors. Furthermore, it was found that the raters paid attention to issues related to source use including: locating source information, citation mechanics, and quality of source use. As integrated tasks have a complex nature, they recommended to use writing professionals in scoring and rating of these tasks.

Research on integrated tasks is critical in highlighting how integration impacts students and their writing performance, how they differ in approaching integrated tasks in comparison to traditional independent writing tasks, as well as in implications for instruction and assessment of integrated writing. Such research is needed to inform teachers and test developers on the impact of using or choosing between the two task types and to help guide development or use of the tasks. Furthermore, as high-stake tests such as IELTS produce high levels of anxiety on the part of learners, any attempts to analyze its tasks, especially the writing section, is important for candidates. Accordingly, it seems vitally important that researchers delve both the test takers performances in accomplishing different writing tasks and the raters' scoring of these tasks. The present study investigated Iranian EFL writing scores in IELTS integrated writing tasks compared to independent tasks. Thus, the following research questions were addressed:

Q1. Is there any significant difference in the writing performance of EFL students in integrated and independent writing tasks?
Q2. Are the raters' performances significantly different in rating independent and integrated IELTS academic writing tasks?

Methodology

Participants

The participants of the present study were 30 Iranian university students. They were majoring in English Language and Literature at Shiraz University, Iran in 2016. The participants were male (46.67%) and female (53.33%) non-native English speaking students whose first language was Persian and were ranged in age from 21 to 25. All the participants were advanced EFL learners and were selected based on their proficiency level and their writing ability.

Raters

Three experienced raters worked on scoring writing samples. All the raters had experience in English language teaching and they had experience in teaching and rating IELTS writing tasks.

They were nonnative English speakers and had Ph.D. degree in TEFL. They rated the full set of essays (60) using the rating rubrics of IELTS integrated and independent writing tasks.

It should be added that after rating all the samples with the interval of one month, the researcher randomly selected 20% of the samples and asked the raters to rerate them. The intra-rater reliability was calculated which was acceptable (r = 0.78).

Scoring rubrics and rater training

The scoring rubrics used in this study were adapted from the IELTS integrated and independent academic writing task rubrics. Both rubrics include a scale ranging from 0 to 9. These rubrics consider 4 dimensions, including "*Task Achievement*" (for integrated task) "*Task Response*" (for independent task), "*Coherence and Cohesion*", "*Lexical Resource*", and "*Grammatical Range and Accuracy*" (Jakeman & McDowell, 2008).

The raters were trained before the initial rating session. Raters were asked to read both scoring rubrics closely. In the training session the raters discussed the scoring rubric; questions and concerns about rating and rubrics were addressed. Then, they were asked to rate one integrated essay and one independent essay at the training session. Upon completing the training, the raters were each asked to rate the full set of 60 essays (30 integrated and 30 independent).

Materials

A number of 60 writing essays were collected from the responses of the participants to the IELTS academic writing Tasks. One of the selected prompts was used to represent the independent category and the other prompt was used with the integrated tasks (McCarter, 2002). In Task 1 (the integrated task) the participants were asked to summarize the information presented in a table in their own words. They should include the most important and the most relevant points. In Task 2 (the independent task) the test takers were presented with a topic to write about. Their answers should be a discursive consideration of the relevant issues. They were asked to read the task carefully and provide a full and relevant answer. They were required to write at least 150 words in 20 minutes on Task 1, and they were asked to write at least 250 words in 40 minutes in answer to task 2; they were penalised if their answers were very short in both tasks. The necessary instructions about how participants should complete writing tasks were presented written for both tasks.

Data collection procedure

The participants were asked to complete the independent and integrated on IELTS writing tasks. In the integrated task, the participants were asked to write a report for a university lecturer describing the information presented in a table about the percentage of the rooms occupied in six hotels during May to September between 1985 and 2000. They should write in an academic or semi-formal/neutral style in the allotted time. In the independent task, they were asked to provide a full and relevant response to the topic "nowadays in countries like Russia some people try to find their matches for marriage through the internet. While some of these relationships have been reported to have happy endings, traditional marriage are more dependent and stable." The students' samples were typed, and their names were replaced by numbers as ID to protect their anonymity.

Data analysis procedure

Descriptive statistical analyses (the mean and standard deviation) as well as paired samples t-tests and one-way repeated ANOVA were conducted on the writing scores given by the

three raters on both tasks (i.e., integrated task and independent task). The paired samples t-tests was run in order to scrutinize if there was any differences between the participants performances while two one-way ANOVA tests were conducted to inspect the difference among raters' performance in rating independent and integrated tasks.

It is worth adding that Microsoft EXCEL was used for data preparation. Further, SPSS was used to conduct both descriptive and inferential statistical analyses. SPSS is a popular data-analysis program used by researchers in social sciences. SPSS can be used for manipulating data, analyzing data, and generating graphs and tables.

Results

Descriptive results

As mentioned, 60 papers on two tasks (integrated and independent writing tasks) were rated by three raters on a 1-9 point scale. Table 1 provides the descriptive statistical results (i.e., the mean and standard deviation of the scores assigned by each rater).

Table 1. *Descriptive statistics for raters' mean scores*

Rater	integrated task		Independent task	
	M	*SD*	*M*	*SD*
1	5.05	1.41	6.45	1.51
2	6.58	.47	6.53	.52
3	6.13	2.13	5.67	1.77
Total	5.92	1.12	6.22	.913

As shown in Table 1, rater 2 and rater 3 assigned higher scores to integrated task than independent task, while rater 1 assigned higher score to independent task which may be due to the participants' performance in these tasks.

Table 1 also shows that the standard deviations of the scores assigned by rater 1 and rater 2 were higher for independent task while for rater 3 it was higher in integrated task. The total scores suggest that the raters were more variant and less consistent in scoring the integrated task.

Paired Samples t-Tests results

The paired samples t-tests were conducted to compare the mean score differences between the integrated and independent tasks scores. The results are presented in Table 2.

Table 2. *Paired Samples Tests results of comparing integrated and independent writing task scores*

	M	t	df	Sig. (2-tailed)
total	.30	1.52	89	.13

**significant at the .05 level

Table 2 reports the mean scores of all raters as the total score. Based on the results, the total score indicated no significant difference between the independent and integrated writing tasks (t (89) = 1.52, n = 30, p < .05).

One-way ANOVA Tests results

Two one-way repeated measure ANOVA tests were conducted to compare scores on integrated and independent writing task assigned by three raters. The related results are presented in Table 3 and Table 4.

Table 3. *ANOVA Test results for raters' performance in rating integrated writing task essays*

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
raters	Pillai's Trace	.618	22.666b	2.000	28.000	.000	.618
	Wilks' Lambda	.382	22.666b	2.000	28.000	.000	.618
	Hotelling's Trace	1.619	22.666b	2.000	28.000	.000	.618
	Roy's Largest Root	1.619	22.666b	2.000	28.000	.000	.618

The results of one-way repeated measure ANOVA test on integrated writing task scores revealed that here was a significant effect for raters, Wilks' Lambda = 0. 38, F (2, 28) = 22.66, p≤.001, multivariate partial eta squared=. 62. Conclusively, there was statistically significant difference in integrated writing task scores across the three raters.

Table 4. *ANOVA Test results for raters' performance in rating independent writing task essays*

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
raters	Pillai's Trace	.190	3.277b	2.000	28.000	.053	.190
	Wilks' Lambda	.810	3.277b	2.000	28.000	.053	.190
	Hotelling's Trace	.234	3.277b	2.000	28.000	.053	.190
	Roy's Largest Root	.234	3.277b	2.000	28.000	.053	.190

According to Table 4, the value for Wilks' Lambda is 0.81, with a probability value of 0.053 and the p value is more than 0.05. Therefore, there was not a significant effect for raters, Wilks' Lambda=0.81, F (2, 28) = 3.277, p≥.05, multivariate partial eta squared= .19.

Discussion

The first research question attempted to determine if there would be significant differences between the test takers' performance in integrated and independent tasks. Although descriptive statistical analysis indicated that the test takers had different performances in these two tasks, the paired samples t-tests specified that this difference was not statistically significant. The findings of the present study are in line with Gebril's (2006) findings indicating a high correlation between the scores on integrated and independent writing tasks. In other words, integrated and independent tasks tap the same construct or relevant aspects of the same construct; that is, academic writing. Consequently, as Gebril (2006) suggested using both tasks together instead of reliance on just one of them would be more beneficial for test takers.

The second research question inspected any differences among raters' performances in rating integrated and independent writing tasks. The obtained results of ANOVA tests specified that the raters had different performances in rating integrated tasks while there was no evidence

of significant change across rating independent tasks samples among the three raters. In writing scoring, raters' variance is inevitable across different writing tasks which may be due to the nature of writing tasks. The integrated task requires more impressionistic judgment than the independent task. The raters in rating the integrated task should judge "if the test takers used the source text effectively or not, if they gave credit to the authors or not, and so on" which is a source of variability among raters (Gebril, 2009, p.523). In comparative studies, different results were reported; Jennings, Fox, Graves, and Shohamy (1999) and Weigle (1999) reported higher and more consistent marks in independent tasks, while Grabe (2001) and Spack (1993) concluded that raters assigned higher scores to integrated writing tasks.

It is worth mentioning that, most of studies did not specifically worked on IELTS academic writing tasks. They mostly compared a certain type of an integrated task with an independent task. Since an integrated task may come in different variations, comparison of the results of such studies need caution, because the agreement or disagreement among the findings may be due to the different variations of the task used not the students or raters' performance. Additionally, the differences mentioned in some studies between integrated and independent tasks in terms of eliciting different behaviors and writing performance (e.g., Guo, 2011) cannot be only judged based on the scores obtained on the two tasks.

Conclusions

The present study focused on IELTS academic writing tasks to see whether different task types have influence on the test takers' performance. The results indicated that test takers may have similar performance on the independent and integrated tasks, in case the raters are experienced and use the same rating rubric.

These results stress the need for instructing EFL learners about the nature of writing tasks and the techniques engaged in different types of writing tasks. For example, teachers can assist students in writing integrated tasks by explicitly teaching them the graphical literacy and empowering them with writing strategies that are helpful in comprehending and describing graphs and tables.

As integrated and independent tasks yielded statistically not different scores, the integrated tasks could be used in second language testing successfully. This result provides useful implications for second language assessment and supports the movement towards new test methods in writing assessment. In fact, by testing writing performance through both integrated and independent writing tasks, a writer's competency in writing skill can be assessed more reliably, and the results can be more accountable. Additionally, this movement could achieve positive washback in academic writing classes where more emphasis would be placed on source-based writing instruction. Moreover, the results can provide important implications for language testing policy makers, teacher preparation institutions/programs as well as EFL teachers and EFL teacher candidates. The results suggested that EFL performance ability raters make it a priority to enhance their professional skill requirement such as language skill assessment.

While the results add important information to the research base regarding writing performance assessment, some problems remain unresolved and should be studied further. The lack of agreement between studies regarding the technical adequacy of holistic and analytic scoring methods still needs to be addressed.

Like all studies, this study also suffered from some limitations; the sample was selected based on their availability and were limited to one specific context. As such, caution needs to be employed in generalizing the findings of the study.

References

Barkaoui, K. (2008). *Effects of scoring method and rater experience on ESL essay rating processes and outcomes*. (Unpublished Doctoral Dissertation). Canada: University of Toronto.

Barkaoui, K., Brooks L., Swain M., Lapkin S. (2013). Test-takers' strategic behaviors in independent and integrated speaking tasks. *Applied Linguistics (Oxford), 34*(3), 304–324.

Brown, A., Iwashita, N., & McNamara, T. (2005). *An examination of rater orientations and test-taker performance on English-for-Academic-Purposes speaking tasks* (TOEFL Monographs Series   MS29). Princeton, NJ: Educational Testing Services.

Brown, J.D., Hudson, T., Norris, J. & Bonk, W.J. (2002). *An investigation of second language task-based performance assessment.* Honolulu, HI: university of Hawaii at Manoa National foreign Language Resource Center.

Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, *8,* 165-191.

Cho, Y., Rijmen, F., & Novak, J. (2013). Investigating the effects of prompt characteristics on the comparability of TOEFL iBT integrated writing tasks. *Language Testing, 30*(4), 513-534.

Clachar, A. (1999). It's not just cognition: The effect of emotion on multiple-level discourse processing in second-language writing. *Language Sciences, 21,* 31–60.

Coffman, W. E. (1971). Essay examinations. In R. L. Thorndike (Ed.), *Educational Measurement*. Washington D. C: American Council on Education.

Cohen, A. (1994). English for academic purposes in Brazil: the use of summary tasks. In C. Hill, & K. Parry (Eds.), *From testing to assessment: English as an international language* (pp. 174-204). London: Longman.

Cumming, A. (1989). Writing expertise and language proficiency. *Language Learning*, *39,* 81- 141.

Cumming, A., Kantor, R., & Powers, D. (2001). *Scoring TOEFL essays and TOEFL 2000 prototype tasks: An investigation into raters' decision making and development of a preliminary analytic framework (TOEFL Monograph Series, Report No: 22).* Princeton, NJ: Educational Testing Service.

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Differences in written discourse in writing-only and reading-to-write prototype tasks for next generation TOEFL. *Assessing Writing, 10*(1), 5-43. doi:10.1016/j.asw.2005.02.001.

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 writing framework: A working paper* (TOEFL Monograph Series, Report No. 18). Princeton, NJ: ETS.

Elder, C., & O'Loughlin, K. (2003). *Investigating the relationship between intensive EAP training and band score gain on IELTS* (IELTS research reports Vol. 4, pp. 207-254). Canberra: IELTS Australia.

Ellis, M., Chong, S., & Choy, Z. (2013). IELTS as an indicator of written proficiency levels: A study of student teachers at the National Institute of Education, Singapore. *International Journal of Educational Research, 60*, 11-18.

Emig, J. (1997). Writing as a mode of learning. In Villanueva, V. (Ed.), *Cross talk in composition theory*. Urbana, IL: National Council of Teachers of English.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *The Canadian Modern Language Review, 58*(4), 599-622.

Evans, N. W., Hartshorn, K. J., McCollum, R. M., & Wolfersberger, M. (2010). Contextualizing corrective feedback in second language writing pedagogy. *Language Teaching Research*, *14*(4), 445–463.

Gebril, A. (2006). *Writing-only and reading-to-write academic writing tasks: A study in generalizability and test method,* (Unpublished doctoral dissertation). The University of Iowa, Iowa City, USA.

Gebril, A. (2009). *Score generalizability in writing assessment: The interface between applied linguistics and psychometrics research.* Saarbrücken, Germany: VDM Verlag Dr. Müller.

Gebril, A. & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. *Spaan Fellow Working Papers in Second or Foreign Language Assessment*, *7*, 47-84.

Gebril, A., & Plakans, L. (2013). Towards a transparent construct of reading-to-write assessment tasks: The interface between discourse features and proficiency. *Language Assessment Quarterly*, *10*(1), 1–19. http://dx.doi.org/10.1080/15434303.2011.642040.

Gebril, A., & Plakans, L. (2014). Assembling validity evidence for assessing academic writing: Rater reactions to integrated task. *Assessing writing*, *21*, 56-73. doi: 10.1016/j.asw.2014.03.002.

Ghalib T., K & Al-Hattami, A., A. (2015). Holistic versus analytic evaluation of EFL writing: A Case Study, *English Language Teaching, 8*(7), 225-236.

Gholami, J., & Alinasab, M. (2016). Iranian EFL Learners' Use of Self-regulatory, Test-wiseness and Discourse Synthesis Strategies in Integrated Writing Tasks. *Social sciences and humanities, 24* (2), 839 – 854

Golder, K., Reeder, K., & Fleming, S. (2011). Determination of appropriate IELTS writing and speaking band scores for admission into two programs at a Canadian post-secondary Polytechnic institution. *The Canadian Journal of Applied Linguistics, 14*(1), 222-250.

Grabe, W. (2001). Reading-writing relations: Theoretical perspectives and instructional practices. In D. Belcher & A. Hirvela (Eds.). *Linking literacies: Perspectives on L2 reading-writing connections* (pp. 15- 47). Ann Arbor, MI: University of Michigan Press.

Green, A. (2007). Washback to learning outcomes: A comparative study of IELTS preparation and university pre-sessional language courses. *Assessment in Education: Principles, Policy & Practice, 14*(1), 75-97.

Guo, L. (2011). *Product and process in TOEFL iBT independent and integrated writing tasks: A validation study* (Unpublished doctoral dissertation). Georgia State University, Georgia, USA.

Guo, L. Crossley, A. S., & McNamara, S. D. (2013). Predicting human judgments of essay quality in both integrated and independent second language writing samples: A comparison study. *Assessing Writing, 18*(3), 218-238. doi: 10.1016/j.asw.2013.05.002.

Hamp-Lyons, L. (1991). Pre-text: Task-related influences on the writer. In L. Hamp-Lyons (Ed.). *Assessing second language writing in academic contexts* (pp. 87-110). Norwood, NJ: Ablex.

Hamp-Lyons, L., & Kroll, B. (1996). Issues in ESL writing assessment: An overview. *College ESL, 6*(1), 52-72.

Hirvela, A., & Du, Q. (2013). Why am I paraphrasing? Undergraduate ESL writers' engagement with source-based academic writing and reading. *Journal of English for Academic Purposes*, *12*(2), 87-98. doi: 10.1016/ j. jeap. 2012.11.005.

Horowitz, D. (1986). Essay examination prompts and the teaching of academic writing. *English for Specific Purposes*, *5*(2), 107–120.

Huang, J. (2011). Generalizability theory as evidence of concerns about fairness in large-scale ESL writing assessments. *TESOL Journal, 2*(4), 423-443. http://dx.doi.org/10.5054/tj.2011.269751

Huang, J., & Foote, C.J. (2010). Grading between lines: what really impacts professors' holistic evaluation of ESL graduate student writing? *Language Assessment Quarterly, 7*(3), 219-233.

Huang, J., & Han, T. (2013). Holistic or analytic – A dilemma for professors to score EFL essays? *Leadership and Policy Quarterly, 2*(1), 1-18.

Hughes, A. (2003). *Testing for language teachers.* Cambridge: Cambridge University Press.

Jakeman, V., & McDowell, C. (2008). *New insights into IELTS*. Cambridge: Cambridge University Press.

Jamieson, J. M., Eignor, D., Grabe, W., & Kunnan, A. J. (2008). Framework for a new TOEFL. In C. A. Chapelle, M. K. Enright, & J. M. Jamieson (Eds.), *Building a validity argument for the test of English as a foreign language* (pp. 145-186). NY: Routledge.

Jennings, M., Fox, J., Graves, B., & Shohamy, E. (1999). The test-takers' choice: An investigation of the effect of topic on language-test performance. *Language Testing*, *16*(4), 426-456.

Knoch, U., & Sitajalabhorn, W. (2013). A closer look at integrated writing tasks: Towards a more focused definition for assessment purposes. *Assessing Writing*, *18*(4), 300-308. doi: 10.1016/j.asw.2013.09.003.

Krapels, A. R. (1990). An overview of second language writing process research. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 37–56). Cambridge: Cambridge University Press.

Lee, Y. W., & Kantor, R. (2005). *Dependability of new ESL writing test scores: Evaluating prototype tasks and alternative rating schemes* (TOEFL Monograph No. MS-31). Princeton, NJ: ETS.

Leki, I. & Carson, J. (1997). Completely different worlds: EAP and the writing experiences of ESL students in university courses. *TESOL Quarterly, 31*, 36-69.

Leki, I., & Carson, J. (1994). Students' perception of EAP writing instruction and writing across disciplines. *TESOL Quarterly*, *28*(1), 81-101. doi: 10.2307/3587199.

McCarter, S. (2002). *Academic writing practice for IELTS*. London: IntelliGene.

McCulloch, S. (2013). Investigating the reading-to- write processes and source use of L2 postgraduate students in real-life academic tasks: An exploratory study. *Journal of English for Academic Purposes*, *12*(2), 136-147. doi: 10.1016/j. jeap.2012.11.009.

Moore, T., & Morton, J. (1999). *Authenticity in the IELTS academic module writing test: A comparative study of Task 2 items and university assignments* (IELTS Research Reports no. 2.). Canberra: IELTS Australia.

Moore, T. & Morton, J. (2005). Dimensions of difference: A comparison of university writing and IELTS writing. *Journal of English for Academic Purposes, 4*(1)*,* 43-66.

Plakans, L. (2007). *Second language writing and reading-to-write assessment tasks: A process study* (Unpublished doctoral dissertation). The University of Iowa, Iowa City: IA.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, *13*(2), 111–129.

Plakans, L. (2009). Discourse synthesis in integrated second language writing assessment. *Language Testing*, *26*(4), 561–587.

Plakans, L., & Gebril, A. (2012). A close investigation of source use in integrated writing tasks. *Assessing Writing Journal*, *17*(1), 18–34. http://dx.doi.org/10.1016/j.asw.2011.09.002

Plakans, L., & Gebril, A. (2013). Using multiple texts in integrated writing assessment: Source use as a predictor of score. *Journal of Second Language Writing, 22*(3), 217-230. doi: 10.1016/j. jslw.2013.02.003.

Raimes, A. (1987). Language proficiency, writing ability, and composing strategies: A study of ESL student writers. *Language Learning, 37,* 439–469.

Read J. (1990). Providing relevant content in an EAP writing test. *English for specific purposes*, *9*(2), 109–121.

Read, J., & Hayes, B. (2003). *The impact of the IELTS test on preparation for academic study in New Zealand* (IELTS research reports Vol. 4, pp. 153-206). Canberra: IELTS Australia.

Sakyi, A. A. (2000). Validation of holistic scoring for ESL writing assessment: How raters evaluate ESL compositions. In A. Kunnan (Ed.), *Fairness and validation in language assessment* (129-152). Cambridge: Cambridge University Press.

Spack, R. (1993). Student meets text, text meets student: Finding a way into academic discourse. In J. Carson & I. Leki (Eds.). *Reading in the composition classroom: Second language perspectives* (pp. 183–196). Boston, MA: Heinle & Heinle.

Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing, 6,* 145-178.

Weigle, S. C. (2002). *Assessing writing.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511732997

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing, 9*, 27-55.

Weigle, C. S., & Parker, K. (2012). Source text borrowing in an integrated reading/writing assessment. *Journal of Second Language Writing*, *21*(2), 118-133. doi: 10.1016/j. jslw.2012.03.004.