# A Comparison of Test-taker Performance on Independent and Integrated L2 Writing Tasks

Özlem Pervan [a]1

*a Maltepe University, Istanbul, Turkey*

**Abstract**

This quantitative, descriptive study investigates the relationship between scores of test-takers on independent and integrated L2 (second language) writing tasks within the context of a foundation university in Turkey. Data were collected from 49 anonymised test scores conducted from independent and integrated L2 writing tasks. The independent task required writing an essay on a general topic, while the integrated task involved responding to prompts based on reading and listening materials. The Wilcoxon signed-rank test was used to compare performance scores, revealing a statistically significant difference between the two tasks ($p < .05$). Specifically, the mean score for the integrated writing task ($M = 70.92$, $SD = 28.32$) was higher than that for the independent task ($M = 66.02$, $SD = 23.96$). Additionally, the Spearman rho test indicated a significant positive correlation between performances on the two tasks ($r = .81$, $p < .01$). These findings suggest that while there is a significant difference in performance levels, there is also a strong positive relationship between the two types of tasks. The results have implications for understanding how different L2 writing tasks affect test-taker performance. Moreover, the results could inform the development of more effective writing assessments and teaching strategies.

*Keywords:* L2 writing tasks; independent writing task; integrated writing task; performance comparison, assessment

## 1.    Introduction

In academic settings, there are two main types of tasks used for L2 writing assessment. One of them is independent L2 writing tasks which require test-takers to compose an essay on a given topic in pre-determined time periods. As cited in Guo, Crossley, & McNamara (2013), many authors such as Camp (1993), Cumming, Kanter, Powers, Santos, & Taylor (2000), and Weigle (2002) claim that independent L2 writing assessment may underrepresent and therefore somehow conceal the test-takers' actual writing proficiency. This underrepresentation may occur due to the risk of decontextualising the writing activity. As cited in Plakans (2008) these

---

[1]  Corresponding author: Özlem Pervan. ORCID ID.: https://orcid.org/0000-0002-6142-0805
    *E-mail address*: ozlempervan@gmail.com

independent writing tasks have also been criticised on issues regarding the test-rater and rating, authenticity of these tasks, and construct validity by Cumming (1997), Lumley (2005) and Weigle (2002).

An alternative to independent L2 writing tasks, is using integrated L2 writing tasks which require test-takers to read and/or listen to source texts before writing an essay. As Jamieson et. al. (2008) suggest, integrated tasks requires the test-taker to use two or more skills, whereas independent tasks are assumed to take speaking or writing as a separate skill. As Jamieson et. al. (2008) suggest, integrated tasks requires the test-taker to use two or more skills, whereas independent tasks are assumed to take speaking or writing as a separate skill.

Instructors in academic settings either prefer to use one of these task types (independent or integrated tasks) for writing assessment, or they use both for different purposes. Questioning whether a relationship exists between the performances of test-takers on these two different types of L2 writing tasks is important to make sure that neither of these tasks underrepresent the L2 writing proficiency of test-takers.

Test of English as a Foreign Language (TOEFL) consists of both independent and integrated writing tasks. The inclusion of integrated writing tasks has caught the attention of many researchers since the late 90's and the early 2000s, and as cited in Cumming et.al. (2005), the arguments and the sources of evidence presented have suggested that including integrated L2 writing tasks in L2 writing assessment for academic purposes provide better measurement of test-takers' L2 writing abilities, require test-designers to design tasks that authentically resemble the writing tasks that the test-takers will face during their academic studies, and so improve the washback effects the tests will have on the process of teaching and learning. Weigle (2004) also states that some scholars consider independent writing tasks to lack authenticity since the test-takers in academic settings are unlikely to face a task where they are going to be required to write an essay on a topic that is previously unseen for them in their content courses. Weigle (2004) also emphasises the fact that academic writing is rarely done in isolation but is almost always done in response to a source text (written or spoken).

The criticisms regarding the use of independent tasks may have caused a shift from relatively more traditional independent, writing-only tasks to integrated, reading-to-write and/or listening-to-write which requires the test-takers to use two or more skills together. However, independent essay writing tasks are still commonly used in English as a Foreign Language (EFL) and English as a Second Language (ESL) contexts. This study brings attention to the issue of whether a performance difference exists between these two different types of tasks within in the context of a foundational university where the mainstream task used for assessing writing is still independent L2 writing. The results of this study may

provide an opportunity to revise the current preferences made for assessing language learners' writing abilities or to prove that what currently is being implemented does not actually underrepresent test-takers' L2 writing proficiency.

This quantitative, descriptive study aims to determine the relationship between test-takers' performances on independent and integrated L2 writing tasks in the context of an English Language Teaching department within a foundation university in Turkey. The null hypothesis of this study is that there is no significant relationship between test-takers' performances on independent and integrated L2 writing tasks.

## 2.    Method

The representative sample for data analysis was selected from 49 anonymised independent and integrated writing task scores collected from the same group of test-takers at a foundation university in Turkey. Independent and integrated-writing tasks of a free sample of TOEFL IBT test were administered. Reliability and validity are declared to be essential aspects of TOEFL scores. The *TOEFL*® research programme follows established guidelines and practices for the development and operational implementation of educational measurements to ensure reliability and validity.

As Alderson (2009) states in his test review of TOEFL IBT, it measures the ability to use and understand English in academic settings among non-native speakers of English. Therefore, TOEFL IBT seems to be a suitable test to measure foreign language learners' written performances on independent and integrated writing tasks used in the context of a Turkish foundation university where English is taught as a foreign language.

In order to see the relationship between test-takers' performances on the integrated writing task and the independent writing task, two data sets conducted from the same group were analysed. Since the test-scores were not obtained from a previously randomised group, The Wilcoxon signed-rank test which is a non-parametric statistical hypothesis test that is used for comparing repeated measurements on a single sample is used. Another reason for the use of the Wilcoxon signed-rank test is the fact that the population cannot be assumed to be normally distributed.

In addition to the comparing of the means of scores achieved on both tests, Spearman rho test, which is the non-parametric equivalent of Pearson's correlation was run in order to determine whether a correlation exists between the two types of L2 writing tasks used.

## 3.     Results

Data were collected from two different types of writing tasks. The analysis was based on scores from 49 instances where both types of tasks were performed by university-level EFL learners as test-takers.

The first aim was to see if there was any difference between the performances of test-takers on two different types of writing tasks: independent writing task (writing an essay on a given general topic) and integrated writing task (writing a respond to a task which requires the test takers to read and listen to materials on the subject they are going to be asked to write about). In order to compare the means of the two data sets obtained from the independent and integrated writing tasks, the Wilcoxon signed-rank test which is the non-parametric equivalent of paired-sample t-test was ran.

Table 1. *Descriptive Statistics*

|  | N | Mean | Std. Deviation | Minimum | Maximum |
|---|---|---|---|---|---|
| Independent | 49 | 66,0204 | 23,95698 | 10,00 | 100,00 |
| Integrated | 49 | 70,9184 | 28,31507 | 5,00 | 100,00 |

It can be seen from Table 1. that the mean for the integrated writing task was slightly higher than the mean for the independent writing task.

Table.2 *Ranks*

|  |  | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| Integrated - Independent | Negative Ranks | 12[a] | 22,88 | 274,50 |
|  | Positive Ranks | 29[b] | 20,22 | 586,50 |
|  | Ties | 8[c] |  |  |
|  | Total | 49 |  |  |

a. Integrated < Independent

b. Integrated > Independent

c. Integrated = Independent

Table 2 shows us that 12 test-takers among the sample received lower grades on the integrated task than they got on the independent task. On the other hand, 29 test-takers seemed to obtain higher scores on the integrated task when compared to their independent task scores. 8 out of 49 test-takers received the same score from both types of tasks.

Table 3. *Wilcoxon Signed Ranks Test Statistics*

*Test Statistics[a]*

|  | Integrated - Independent |
|---|---|
| Z | -2,028[b] |
| Asymp. Sig. (2-tailed) | ,043 |

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

The hypotheses of this study was that there was no significant difference and/or relationship between test-takers' performances on independent and integrated L2 writing tasks. However, the analysis showed us that the performances were significantly different ($p < .05$). Since the p value turned out to be smaller than .05, based on the analysis, we reject the null hypotheses and accept that there is a significant difference between the test-takers' performances on independent and integrated L2 writing tasks.

In order to see whether a correlation existed between the performances on independent and integrated L2 writing tasks, the Spearman rho test, which is the non-parametric equivalent of Pearson's correlation was administered.

Table 4. *Correlations Matrix of Variables*

| *Correlations Matrix of Variables* (N=49) | | Independent | Integrated |
|---|---|---|---|
| Spearman's rho | Independent | 1 | .81** |
| | Integrated | | 1 |

**. Correlation is significant at the 0.01 level (2-tailed).

As can be seen on Table 4., a significant positive correlation was found between the independent and the integrated L2 writing tasks. Therefore, here the null hypotheses of "there is no significant relationship between test-takers' performances on independent and integrated L2 writing tasks" is rejected and it is accepted that there is a significant and positive correlation between the test-takers' scores and therefore performances on independent and integrated L2 writing tasks.

## 4.    Discussion

This study was conducted in order to determine whether a difference and/or a relationship existed between test-takers' performances on independent and integrated L2 writing tasks within the context of a foundation university in Turkey. Being able to determine whether such relationship existed was considered important by the researcher as integrated L2 writing tasks were considered to be more meaningful and authentic for the test-takers in the sense that they resembled real writing tasks where university students would be expected to write responses to the materials they have read and/or listen to instead of being required to ask essays on given general topics.

Since the early 1990's many authors such as Camp (1993); Cumming, Kanter, Power, Santos & Taylor (2000); Weigle (2002); Guo, Crossley & McNamara (2013) claim that independent writing tasks underrepresents test-takers' L2 writing proficiencies. In line with these claims, the results of this study shows that same test-takers showed significantly higher performances in the integrated L2 writing task compared to independent L2 writing task. The significant positive correlation found between the performances of the same test-takers in two different tasks shows us that test-takers who performs relatively well on one of the tasks is more likely to perform relatively well on the other task. However, the significant difference between the performances still cannot be ignored and should be taken into consideration.

## 5.    Conclusions

Both independent and integrated L2 writing tasks are designed to measure the same construct, which is L2 writing ability. The significant difference found between the performances of the same test-takers on the two different types of tasks shows that although they measure the same construct independent L2 writing tasks significantly underrepresents test-takers' L2 writing abilities as suggested in the previous literature.

The decision of using which type of L2 writing tasks belongs to institutions and instructors. However, this decision might not always have a rationale behind it except for the fact that independent L2 writing tasks have been used for a much longer time than the relatively recent integrated L2 writing tasks. This study might contribute to the understanding of current EFL (English as a Foreign Language) practitioners in the sense that it points out the fact that integrated L2 writing tasks do not only have a significant positive correlation with independent L2 writing tasks, but also they might provide a more accurate representation of the actual L2 writing abilities of EFL learners. Moreover, the findings can also help other instructors in similar contexts to improve their teaching and learning environment.

## References

Alderson, J. C. (2009). Test review: Test of English as a foreign language: Internet-based test (TOEFL iBT®). *Language Testing*, *26*(4), 621-631. https://doi.org/10.1177/0265532209346371

Camp, R. (1993). Changing the model for the direct assessment of writing. In *Validating Holistic Scoring for Writing Assessment: Theoretical and Empirical Foundations* (pp. 45–78). Hampton Press.

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

Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). TOEFL 2000 writing framework. In *TOEFL-MS-18*. Educational Testing Service Princeton, NJ.

Guo, L., Crossley, S. A., & McNamara, D. S. (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. https://doi.org/10.1016/j.asw.2013.05.002

Jamieson, J. M., Eignor, D., Grabe, W., & Kunnan, A. J. (2008). Frameworks for a new TOEFL. *Building a validity argument for the Test of English as a Foreign Language*, 55-95.

Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, *13*(2), 111-129. https://doi.org/10.1016/j.asw.2008.07.001

Weigle, S. C. (2002). *Assessing writing*. Ernst Klett Sprachen.

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing writing*, *9*(1), 27-55. https://doi.org/10.1016/j.asw.2004.01.002