Sino-US English Teaching, August 2024, Vol. 21, No. 8, 354-371
doi:10.17265/1539-8072/2024.08.002

# Analyzing the Essay Writing Patterns of Iranian Intermediate-Level Korean Language Learners: A Study Based on the TOPIK Exam

Vahid Reza Nasrollahi
Seoul National University, Seoul, South Korea

This study examines the writing abilities of Iranian intermediate Korean learners, specifically their performance in the compositional writing section of the Test of Proficiency in Korean (TOPIK). By utilizing the many FACETS-Rasch Model, we meticulously analyzed nine writing samples from the 52nd TOPIK. These samples were evaluated using a modified rubric ranging from 0 to 3 based on predefined criteria for written composition. The results underscored that the sections on "appropriateness of spacing and spelling," "relevance of vocabulary," and "content diversity" presented the most significant challenges for the learners. On the other hand, "the quantity of writing" emerged as the least challenging aspect. These findings reveal substantial disparities in various aspects of writing proficiency among learners. The study not only pinpoints issue areas in writing skills, but also underscores the necessity of customized teaching strategies within the TOPIK framework to address these weaknesses. Consequently, it offers valuable insights that could bolster the effectiveness of teaching writing to Korean language learners. The findings of this study are not only significant for the field of language education, but also contribute to a deeper understanding of the challenges faced by intermediate learners and provide a roadmap for improving language instruction.

*Keywords:* TOPIK writing evaluation, Iranian intermediate Korean learners, multinational research analysis, writing challenges, language education strategies

## Introduction

Today, the Korean Wave continues to exert a profound cultural influence worldwide, with Iran emerging as a notable market in the Middle East and North Africa regions. This phenomenon underscores the deep-rooted historical and cultural ties between Korea and Iran, dating back 1200 years to the Silk Road era, as evidenced in traditional texts like "Samguk Yusa" (Memorabilia of the Three Kingdoms), where Cheoyong, possibly a Persian merchant, highlights early cultural exchanges (Choi, 2009). The academic interactions between South Korea and Iran have surged, with the number of Iranian students in Korean Higher Education growing from 12,388 in 2008 to 48,153 in 2023 (The Ministry of Education, 2023). Despite this growth, challenges persist, notably in the academic completion rates of international students, including Iranian (Ahn, 2012). The Test of Proficiency in Korean (TOPIK), established in 1997 and managed by the National Institute for International Education (NIIED),

Vahid Reza Nasrollahi, Ph.D. candidate, Department of Korean Language Education—Teaching Korean as a Second Foreign Language, Seoul National University, Seoul, South Korea.

plays a crucial role in assessing and supporting the academic skills of international students. Conducted multiple times annually in 53 other countries, TOPIK evaluates proficiency in vocabulary and grammar, writing, listening, and reading. The writing section requires advanced analytical skills, particularly in argumentative writing, which has been identified as a significant challenge for students (Yan & Cheng, 2015). In 2014, TOPIK's structure was revised to emphasize argumentative writing, which now accounts for 50 out of 100 points in the writing section. This change underscores the importance of advanced analytical skills and logical writing (Kim, Lee, Pyun, & Byon, 2022). Studies revealed that argumentative writing is one of the most challenging tasks, often resulting in lower scores for advanced learners (Batjargal & Enkhtuya, 2017). For Iranian beginners, challenges in pronunciation and writing order are notable, paralleling challenges in argumentative writing (Nasrollahi, 2020).

Writing is frequently the most challenging skill for Korean language learners (Kim, 2000) often causing significant stress and pressure among learners (Lee, 2003). For advanced learners, particularly those with academic purposes, writing is the skill that causes the most stress. Advanced learners, particularly those pursuing academic purposes, are evaluated on their ability to produce logical and persuasive written work. Consequently, mastering argumentative writing is crucial for academic success in Korean Higher Education (Jeong, 2015). Research on writing education in South Korea has predominantly focused on text structure (Min, 2001, 2004a, 2004b; Yang, 2008), with limited exploration of genre-specific characteristics of argumentative writing. Studies have examined grammatical expression and writer intent in argumentative texts (Kim, 2009), proposed teaching and learning strategies (Nguyen, 2015), and identified genre-specific features (Nam, 2011). These studies, primarily conducted within academic discourse, have concentrated on the methodology linking reading and writing. They highlight the necessity of examining not only the structural components of written texts, but also the distinct cultural aspects reflected in the form and content of learners' argumentative writing. This underscores the importance of developing genre-based writing education, supported by research on the organization of content and the linguistic characteristics of specific genres (Lee, 2010).

While research on Korean argumentative writing specifically for Iranian learners is limited, extensive international discourse has focused on English argumentative writing for these learners. The research primarily addresses Iranian English as a foreign language (EFL) learners, aiming to enhance their argumentative writing skills. For instance, Khodabandeh et al. (2014) demonstrated that training Iranian learners in English rhetorical patterns significantly improved their writing abilities. Nimehchisalem et al. (2015) found that Iranian EFL learners excelled in "content" and "organization," showed general proficiency in "task performance," "vocabulary," and "writing style," average in "grammar," and weakest in "mechanical skills," which includes all mentioned categories. The population identified that the most frequent errors among these learners involved tense, cohesion, consistency, and vocabulary. Salmani Nodoushan (2018) categorized most writing errors as structural and cognitive. Additionally, Derakhshan and Karimian Shirejini (2020) noted that Persian language interference negatively affected the learners' mastery of conversational English, syntax, and rhetorical structures. Conversely, Marashi (2020) highlighted that guidelines on grammatical and lexical cohesion were beneficial for Iranian learners in comprehending the rationale behind their writing. These results align with the observation that foreign learners, particularly Iranian students, face significant challenges with argumentative writing in Korean, especially regarding cohesion, consistency, and vocabulary (Shin, 2017). It confirms that research in Korean language learning has primarily concentrated on structural elements, highlighting these specific challenges. This study provides insights into Korean argumentative writing instruction for intermediate Iranian learners by

emphasizing the importance of foundational writing structures and incorporating Korean-specific genre and rhetorical patterns into the training.

Building upon previous discussions and research, this study addresses the challenges faced by Iranian intermediate Korean learners in the TOPIK writing section. The primary objective is to enhance their argumentative writing skills by identifying specific challenges that hinder their performance. By examining the characteristics of the content and form of these learners' writing, the study aims to deepen understanding of the genre knowledge required for argumentative texts. To achieve this, the study will analyze answer sheets from the 52nd TOPIK essay writing examination post-revision to identify specific errors and their causes (Ma, 2019; Lee, 2017). Additionally, it will review earlier works on writing evaluation criteria (Dashdorj & Sodnomdorj, 2020; Song, 2020; Park, 2016; Oh, 2015; Park, 2009; Kim, 2004; Kim, 2005; Seo, 2003, 2008) focusing on the three critical components of writing: content, organization, and expression. This comprehensive approach aims to use content and evaluation methods, improve argumentative writing techniques and content, and develop targeted strategies to strengthen specific areas of weakness identified in the study.

The study is structured as follows: Section 2 details the research methodology, including data collection, scoring procedures, and rater profiles. Section 3 presents the analysis results, evaluating data appropriateness, measurement profile distributions, and profiles of examinees and raters. And Section 4 concludes with a discussion of the findings and their broader implications for future research.

## Research Methodology

This section discusses the methods used to evaluate Iranian intermediate Korean learners' writing abilities, emphasizing their performance on the TOPIK exam's compositional writing section. The study used a mixed-methods approach, integrating quantitative and qualitative data collection and analysis to provide a comprehensive understanding of the learners' writing abilities and limitations.

### Data Collection

**Participants Selection.** The study's initial step involves rigorous participant selection, focusing on Iranian students at the intermediate level enrolling at several Korean universities. Twenty students were interviewed as part of a preliminary evaluation to determine their degree of willingness and readiness to take the TOPIK writing exam. The selection criteria included their current academic level and competency in the Korean language. Nine students from the initial group agreed to participate, providing written agreement while upholding ethical norms and voluntary participation.

**Test Preparation.** To ensure the reliability and validity of the data, the TOPIK writing test was conducted under regulated conditions:

*Controlled environment*: The test was conducted in a peaceful and undisturbed environment to ensure uniformity among all participants.

*Time constraint*: Participants were required to complete the writing activity within a strict time limit of 30 minutes. This accurately reflects the conditions of the real TOPIK exam, emphasizing the significance of their assignment.

*Prohibition of aid*: To maintain the integrity of the test, participants were prohibited from using electronic devices, cell phones, dictionaries, or personal notes.

*Prohibition of prior knowledge*: Discussions regarding the subject of previous exam questions were strictly forbidden to prevent potential prejudice.

**Selection of Writing Samples.** After the examination, the answer sheets were carefully examined. Only answer sheets that were totally completed and had a character count between 600 and 700 were chosen for study. This criterion guaranteed that the instances offered a strong and reliable dataset for thoroughly assessing the trainees' writing abilities.

**Qualitative Content Analysis.** The study's qualitative component specifically examined the responses to the 54th writing assignment of the 52nd TOPIK. The participants were tasked with expressing their ideas on a specific topic, which allowed for an in-depth evaluation of their capability to write persuasive arguments (see Table 1).

Table 1

*Overview of Writing Assignment*

| Korean | English |
|---|---|
| "우리는 살면서 서로의 생각이 달라 갈등을 겪는 경우가 많다. 이러한 갈등은 의사소통이 부족해서 생기는 경우가 대부분이다. 의사소통은 서로의 관계를 유지하고 발전시키는 데 중요한 요인이 된다. '의사소통의 중요성과 방법'에 대해 아래의 내용을 중심으로 자신의 생각을 쓰라."<br>의사소통은 왜 중요한가?<br>의사소통이 잘 이루어지지 않은 이유는 무엇인가?<br>의사소통을 원활하게 하는 방법은 무엇인가? | "Conflicts in our daily lives frequently arise when people hold opposing perspectives. Inadequate communication is frequently the root cause of many confrontations. Effective communication is essential for maintaining and strengthening relationships. Discuss 'the relevance and practices of communication,' emphasizing the following issues."<br>Why is it so important to communicate?<br>What causes ineffective communication?<br>What approaches increase communication effectiveness? |

**Analytical Framework**

**FACET Software Utilization.** The study used the FACETS program (Minifac 3.83.6) to analyze the gathered research data. FACET is a valuable tool for probabilistically estimating examinees' inherent language performance abilities using evaluation results presented as multifaceted scores, such as those from oral tests or writing tasks. It is widely employed in research involving multifaceted Rasch models. It is notable for its effectiveness in many facets of Rasch measurement, which provides a complete study of complex performance evaluations. It allows for a detailed study of numerous facets, such as raters, tasks, and test takers, by isolating and evaluating each facet's contribution to the final performance score. The software enables users to control variables, such as the influence of raters in the measurement model. FACET is adaptable, accepting aspects that can be randomly chosen or fixed, and it successfully manages variability and interactions among multiple facets, offering a thorough understanding of performance assessments (Linacre, 1998; 1996). In this study, FACET was used to evaluate the following aspects:

1. Raters: To evaluate the extent and reliability of each rater's severity and consistency;
2. Purpose: To assess the level of difficulty for each writing activity or rubric criterion;
3. Test-takers: To assess the proficiency levels of the learners.

This holistic approach, provided by the FACET program, is critical for resolving any biases and inconsistencies. It results in a more accurate and equal evaluation of the learners' writing proficiency, which improves the assessment process's credibility and dependability (Engelhard Jr, 2013).

**Multinomial Rasch Model.** The ordinal data obtained from the rubric scores was analyzed by using the multinomial Rasch model in the FACETS software. This model is particularly appropriate for managing ordered categorical data, making it perfect for analyzing performance-based evaluations like writing tasks. The

multinomial Rasch model calculates the likelihood of achieving a specific score by considering the level of difficulty of the task and the skill level of the test taker. This enables a comprehensive assessment of writing capability.

The multinomial Rasch model employed in this study includes the following prominent characteristics:

*Difficulty estimation*: The model assesses the level of difficulty for each rubric criterion, offering insights into the specific components of the writing task that are more demanding for learners.

*Ability estimation*: The model evaluates the proficiency levels of individual learners, providing an in-depth assessment of their writing skills based on several criteria.

*Interaction analysis*: The model facilitates the identification of potential biases and inconsistencies in the scoring process by analyzing the interactions among raters, tasks, and examinees.

The multinomial Rasch model is a highly effective tool for evaluating and enhancing performance-based evaluations due to its capability to handle numerous aspects and ordered categorical data. The tool offers comprehensive diagnostic information that can be utilized to improve teaching methods and assessment design (Engelhard Jr, 2013).

**Raters**

Three individuals, including the author, were assigned the task of evaluating the writing abilities of Iranian intermediate Korean learners who were taking the TOPIK test (see Table 2).

According to Table 2, the group comprised two specialized Korean instructors who taught Korean and TOPIK at an official Korean language education center for foreigners. All Korean raters have a minimum of six years of experience teaching and evaluating Korean language proficiency.

Table 2

*Raters' Information*

| Rater | Gender | Nationality | Major | Education level | Korean language /TOPIK teaching experience |
|---|---|---|---|---|---|
| SY | Female | Korean | Korean language education | Ph.D. candidate | Six years |
| JU | Female | Korean | | | Six years |
| DR | Male | Iranian | | | Two years |

**Accuracy and consistency of scoring.** To ensure the accuracy and consistency of the scoring, several measures were implemented:

*Calibration sessions were carried out to synchronize the raters' scoring criteria.* The sessions consisted of evaluating a series of standard essays and addressing any inconsistencies in the scores to establish a unanimous understanding of how the criteria should be applied. This method is crucial for reducing differences between raters and ensuring that all raters consistently understand the scoring criteria (Engelhard Jr, 2013).

*Standardization of scoring criteria.* The rubric's criteria were carefully reviewed and clearly described to the raters to ensure a consistent understanding. This step included an in-depth discussion on the definitions and expectations for each level of scoring within the rubric, ensuring that all raters had a clear and uniform interpretation of the criteria (McNamara, 1996).

*Practice scoring: Raters participated in practice scoring sessions using sample essays.* These sessions enabled them to apply the rubric to a variety of responses and provided an opportunity to refine their scoring

approach based on feedback and group discussions. Practice scoring helps to identify potential biases and areas where raters may need additional guidance (Weigle, 2002).

Consistency checks were often conducted throughout the scoring process to ensure accuracy and uniformity. These checks involved comparing scores assigned by different raters to the same essays to identify any significant discrepancies. When divergences were found, further discussions were held to resolve them and ensure consistent application of the scoring criteria (Lumley & McNamara, 1995).

*Feedback and adjustment.* The raters received continued feedback throughout the scoring process. This feedback focused on maintaining consistency and accuracy in scoring. Based on the feedback, adjustments were made as needed to improve the reliability of the scoring process (Knoch, 2009).

Table 3

*Categories and Evaluation Content for TOPIK Essay Tasks*

| Item | Category | Evaluation content |
|---|---|---|
| 54 | Content and Task Achievement (12 points) | Did the response faithfully execute the given task? |
| | | Is it composed of content related to the topic? |
| | | Were the given contents expressed diversely and richly? |
| | Structural Development (12 points) | Is the composition logical and clear? |
| | | Is paragraph structuring well executed and appropriate to the content? |
| | | Are discourse markers used to connect the development of arguments logically and organizationally? |
| | Language Use (26 points) | Were grammar and vocabulary chosen appropriately and used diversely and richly? |
| | | Are grammar, vocabulary, and spelling used accurately? |
| | | Is the writing appropriately formal according to the purpose and function of the text? |

**Source:** Compiled by referring to the Test of Proficiency in Korean website http://topik.go.kr.

Table 4

*Variables Named By Evaluation Content*

| Evaluation Category | Evaluation Content | Variable Name |
|---|---|---|
| Content and Task Achievement | Did the response faithfully execute the given task? | V1_1 |
| | Is it composed of content related to the topic? | V1_2 |
| | Were the given contents expressed diversely and richly? | V1_3 |
| Structural Development | 2.1. Is the composition logical and clear? | V2_1 |
| | 2.2. Is paragraph structuring well executed and appropriate to the content? | V2_2 |
| | 2.3. Are discourse markers used to connect the development of arguments logically and organizationally? | V2_3 |
| Language Use | 3.1. Were appropriate vocabulary choices made? | V3_1 |
| | 3.2. Was diverse vocabulary used? | V3_2 |
| | 3.3. Was appropriate grammar used? | V3_3 |
| | 3.4. Was diverse grammar used? | V3_4 |
| | 3.5. Were correct spacing and spelling used? | V3_5 |
| | 3.6. Was the quantity of writing appropriate? | V3_6 |
| | 3.7. Was the formal language used instead of colloquial language? | V3_7 |
| | 3.8. Was the manuscript paper format well considered? | V3_8 |

**Scoring Procedure.** The essay task was assessed using an analytical evaluation method. This approach classifies the assessment into three primary domains: the quality of the content and task completion, the progression of the essay's structure, and proficiency in language use. To facilitate the analysis of the measured

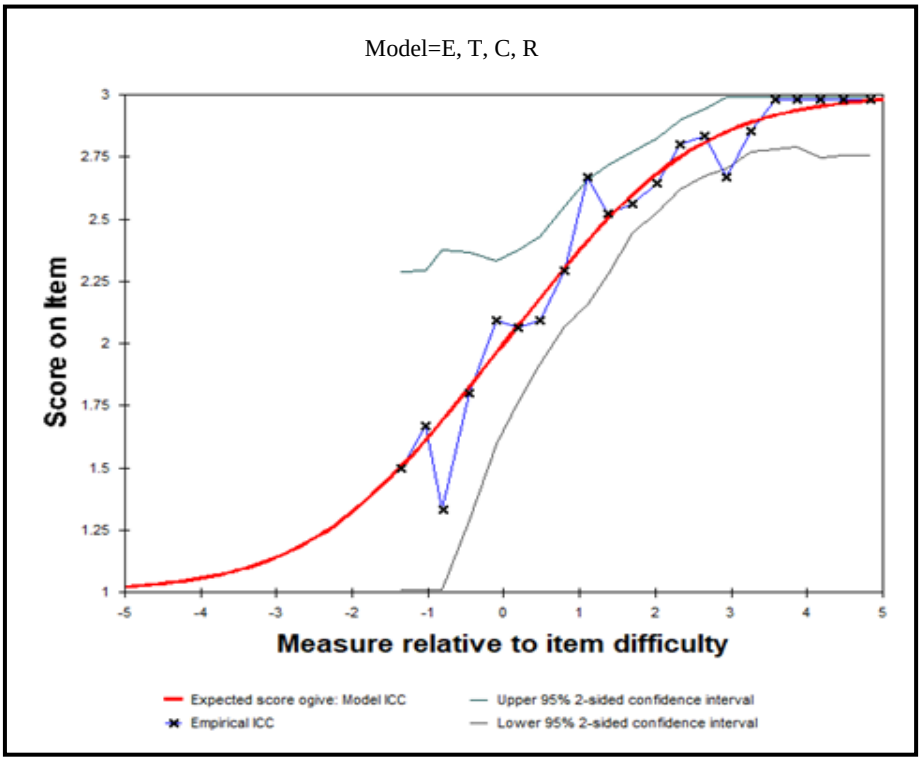
data, the evaluation content was converted into English variables. Table 3 provides a breakdown of each main category into subcategories.

The evaluation information in Table 3 has been modified to consider the need for explicit criteria for the performance scale in each question. To improve clarity and accuracy, the following Dashdorj and Sodnomdorj (2020), Park (2016), Song (2020), and Oh (2015) used a three-point scale ranging from 0 to 3 points were used to revise and rephrase the questions in the evaluation material. This modification preserves the sub-items that were initially included in the vocabulary use category, but it divides vocabulary, grammar, and spelling into separate sub-categories. As a result, the total number of items has increased without introducing any new categories. The scale is calibrated such that a value of 1 corresponds to the minimum standard, while a value of 3 corresponds to the maximum standard (see Appendix). Therefore, to facilitate the analysis of the measured data, the evaluation items were converted into English variables. Table 4 lists the variable names by evaluation content.

## Analysis Results

### Fit of Scoring Data

Figure 1 illustrates the Item Characteristic Curve (ICC) analysis, which shows the relationship between the measure's relative to item difficulty and the item score.




E: examinee facet, T: task facet, C: criteria factor, R: rater facet

*Figure 1.* Item characteristics curve analysis.

The red curve represents the model's expected score ogive[1]. Model ICC, which displays predicted scores based on the examinees' abilities and the item's complexity. The empirical item characteristic curve, symbolized

[1]. The frequency distribution graph of a series.

by black crosses connected by a blue line, reflects the observed scores for each level of difficulty. The green and grey lines indicate the upper and lower 95% two-sided confidence intervals, respectively. These intervals represent the range in which 95% of recorded scores are predicted to fall. The empirical ICC is often consistent with the expected score ogive, indicating that the model's predictions closely match the observed facts. Most observed scores lie between the upper and lower confidence ranges, indicating a strong fit for the model. The graph's horizontal axis extends from -5 to 5, showing the range of item difficulty, and the vertical axis ranges from 1 to 3, indicating the possible scores for each item. The trend demonstrates that as the difficulty measure grows, so does the item's score, implying that higher-ability examinees perform better on more difficult items. The close alignment of the empirical ICC with the anticipated score ogive, as well as the inclusion of the majority of observed scores within the confidence intervals, indicate a good fit between the Rasch model and the actual data. This confirms that the Rasch model is acceptable for assessing the writing task data in this study, the reliability and validity of the scoring procedure, and the accuracy of the item difficulty assessments.

**Analysis of Measurement Facet Distributions**

Figure 2 illustrates the distribution of four measurement facets- examinee, rater, task, and criteria- on a common logit scale.

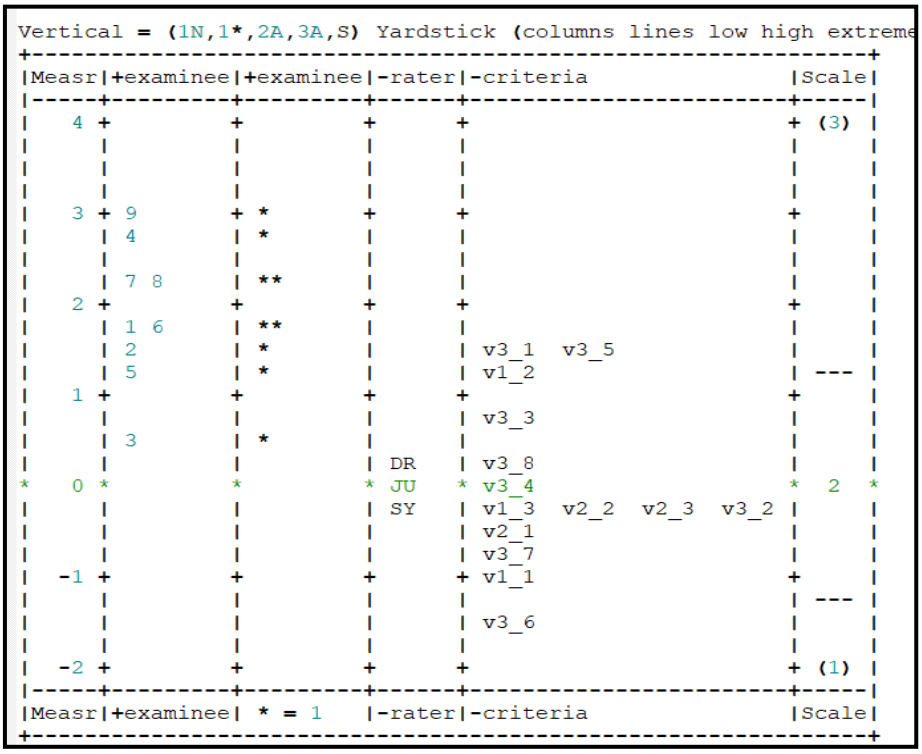



*Figure 2.* Distribution analysis of measurement facets.

In Figure 2, the vertical axis, labeled as "Measr," represents the logit measure, indicating the levels of ability or difficulty. In the examinee facet, participants are denoted by ID and asterisks (*), with higher positions on the scale reflecting higher ability levels, while lower positions indicate lower performance levels. The rater facet is represented by the acronyms DR, JU, and SY, whose locations indicate the severity of their scoring. The rater, identified as DR, is positioned higher on the scale, indicating a tougher scoring approach than JU and SY, who

are put lower and so demonstrate lower scoring tendencies. The criterion facet displays the difficulty levels of the fourteen evaluated tasks, with higher spots indicating more complexity. The tasks relating to the appropriateness of word use (V3_1) and the precision of spacing and spelling (V3_5) were identified as the most difficult for the examinees. In contrast, the assignment involving the quantity of writing (V3_6) is seen as the least demanding.

**Analysis of Examinee Facets**

Table 4 shows that the fit and reliability of the examinee aspects were assessed using the data.

According to the analysis, fit indices are divided into Infit Mean Square (Infit Mnsq) and Outfit Mean Square (Outfit Mnsq). The standard for these indices is 1.0 logits, with a range of 0.5 to 1.5 logits being considered indicative of appropriate learner responses. Indices greater than 1.5 logits are evaluated as misfit models, suggesting that the response patterns are inconsistent with the model, while indices less than 0.5 logits indicate overfit learners, implying overly predictable responses (Dashdorj & Sodnomdorg, 2020; Kim, 2015; Shin & Seol, 2005; Lee, 2012; Choi, 2016).

Table 4

*Analysis Results of Examinee Facets*

analysis of Davis's Article 2021-12-04 오후 4:58:07
Table 7 ' ^ Examinee Measurement Report (arranged by 1N).

| Total Score | Total Count | Obsvd Average | Fair(M) Average | + Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Correlation PtMea | PtExp | N examinee |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 108 | 42 | 2.57 | 2.62 | 1.76 | .30 | .96 | -.1 | .94 | -.1 | 1.07 | .51 | .46 | 1 01 |
| 105 | 42 | 2.50 | 2.54 | 1.50 | .29 | .93 | -.2 | 1.07 | .3 | .96 | .34 | .48 | 2 02 |
| 91 | 42 | 2.17 | 2.18 | .46 | .26 | 1.42 | 2.0 | 1.38 | 1.8 | .42 | .51 | .50 | 3 03 |
| 117 | 42 | 2.79 | 2.83 | 2.79 | .39 | 1.01 | .1 | .98 | .0 | .96 | .28 | .38 | 4 04 |
| 103 | 42 | 2.45 | 2.49 | 1.33 | .28 | .74 | -1.3 | .73 | -1.3 | 1.30 | .51 | .48 | 5 05 |
| 109 | 42 | 2.60 | 2.64 | 1.86 | .31 | 1.10 | .5 | 1.13 | .5 | .91 | .42 | .46 | 6 06 |
| 112 | 42 | 2.67 | 2.71 | 2.16 | .33 | .94 | -.1 | 1.15 | .5 | .93 | .28 | .44 | 7 07 |
| 112 | 42 | 2.67 | 2.71 | 2.16 | .33 | .90 | -.3 | .81 | -.5 | 1.12 | .51 | .44 | 8 08 |
| 119 | 42 | 2.83 | 2.87 | 3.12 | .42 | .69 | -.9 | .43 | -1.3 | 1.24 | .60 | .35 | 9 09 |
| 108.4 | 42.0 | 2.58 | 2.62 | 1.90 | .32 | .97 | -.1 | .96 | .0 | | .44 | | Mean (Count: 9) |
| 7.9 | .0 | .19 | .20 | .74 | .05 | .20 | .9 | .26 | 1.0 | | .11 | | S.D. (Population) |
| 8.3 | .0 | .20 | .21 | .79 | .05 | .21 | 1.0 | .27 | 1.0 | | .12 | | S.D. (Sample) |

Model, Populn: RMSE .33 Adj (True) S.D. .67 Separation 2.04 Strata 3.05 Reliability .81
Model, Sample: RMSE .33 Adj (True) S.D. .72 Separation 2.19 Strata 3.25 Reliability .83
Model, Fixed (all same) chi-squared: 47.9 d.f.: 8 significance (probability): .00
Model, Random (normal) chi-squared: 6.9 d.f.: 7 significance (probability): .44

Based on these criteria, the analysis revealed that all nine Iranian learners have Infit and Outfit Mnsq values between 0.5 and 1.5 logits, indicating that there are neither misfit nor overfit learners among the participants. The mean Infit Mnsq is roughly 1.04, while the mean Outfit Mnsq is approximately 1.01, indicating the appropriateness of the responses. Furthermore, the examinee separation reliability was determined to be 0.83, as evidenced by a separation index of 2.19 and strata of 3.25. This high level of reliability indicates that the item difficulty distribution is well-suited to explaining the variance in examinee abilities, resulting in a reliable measurement of learners' performance.

**Analysis of Rater Facets**

Table 5 presents the analysis of rater facets, which measures the severity of each rater's scoring.

The data indicate that Rater 3 (DR) has the highest severity score at 0.27. This is consistent with the observed average score (Obsvd Average), where Rater 3's average score of 2.52 reflects the most stringent scoring among

the raters. The analysis of both Infit and Outfit Mean Square (Mnsq) values shows that all raters are within the acceptable range of 0.5 to 1.5, centered around the criterion value of 1.0. This indicates that all raters demonstrated internal consistency in their scoring. Additionally, the exact observed agreement percentage (Exact Obs %) and the expected agreement percentage (Exp %) estimated by the model were compared to determine whether the raters scored independently. The results reveal that all three raters exhibited a level of agreement that meets or exceeds the expected values, indicating a high degree of scoring consistency among the raters.

Table 5

*Analysis Results of Rater Facets*

analysis of Davis's Article 2021-12-04 오후 4:58:07
Table 7. ater Measurement Report (arranged by MN).

| Total Score | Total Count | Obsvd Average | Fair(M) Average | Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Correlation PtMea | Correlation PtExp | Exact Agree. Obs % | Exact Agree. Exp % | N rater |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 332 | 126 | 2.63 | 2.71 | -.22 | .19 | .74 | -2.0 | .69 | -1.7 | 1.26 | .62 | .50 | 75.8 | 60.1 | 2 SY |
| 327 | 126 | 2.60 | 2.67 | -.05 | .18 | 1.45 | 3.0 | 1.38 | 2.0 | .57 | .37 | .52 | 63.1 | 59.8 | 1 JU |
| 317 | 126 | 2.52 | 2.58 | .27 | .17 | .81 | -1.5 | .80 | -1.3 | 1.19 | .58 | .53 | 73.8 | 58.5 | 3 DR |
| 325.3 | 126.0 | 2.58 | 2.65 | .00 | .18 | 1.00 | -.2 | .96 | -.4 | | .52 | | | | Mean (Count: 3) |
| 6.2 | .0 | .05 | .05 | .20 | .01 | .32 | 2.3 | .30 | 1.7 | | .11 | | | | S.D. (Population) |
| 7.6 | .0 | .06 | .06 | .25 | .01 | .39 | 2.8 | .37 | 2.1 | | .14 | | | | S.D. (Sample) |

Model, Populn: RMSE .18 Adj (True) S.D. .09 Separation .49 Strata .99 Reliability (not inter-rater) .19
Model, Sample: RMSE .18 Adj (True) S.D. .17 Separation .93 Strata 1.57 Reliability (not inter-rater) .46
Model, Fixed (all same) chi-squared: 3.8 d.f.: 2 significance (probability): .15
Model, Random (normal) chi-squared: 1.3 d.f.: 1 significance (probability): .25
Inter-Rater agreement opportunities: 378 Exact agreements: 268 = 70.9% Expected: 224.7 = 59.4%

**Analysis of evaluation scale Facets**

Table 6 presents the analysis of the evaluation scale facets, which examines the difficulty and fit levels of the evaluation items.

Table 6

*Analysis Results of Evaluation Scale Facets*

analysis of Davis's Article 2021-12-04 오후 4:58:07
Table 7.3.1 criteria Measurement Report (arranged by MN).

| Total Score | Total Count | Obsvd Average | Fair(M) Average | Measure | Model S.E. | Infit MnSq | Infit ZStd | Outfit MnSq | Outfit ZStd | Estim. Discrm | Correlation PtMea | Correlation PtExp | Nu criteria |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 78 | 27 | 2.89 | 2.91 | -1.59 | .62 | .90 | .0 | .62 | -.4 | 1.08 | .37 | .25 | 12 v3_6 |
| 76 | 27 | 2.81 | 2.84 | -.98 | .50 | 1.22 | .6 | 1.31 | .7 | .78 | -.05 | .31 | 1 v1_1 |
| 75 | 27 | 2.78 | 2.81 | -.75 | .47 | 1.19 | .6 | 1.16 | .4 | .81 | .02 | .33 | 13 v3_7 |
| 74 | 27 | 2.74 | 2.77 | -.54 | .44 | .81 | -.5 | .79 | -.4 | 1.14 | .42 | .34 | 4 v2_1 |
| 73 | 27 | 2.70 | 2.74 | -.36 | .42 | .74 | -.8 | .61 | -1.1 | 1.27 | .55 | .36 | 3 v1_3 |
| 73 | 27 | 2.70 | 2.74 | -.36 | .42 | .98 | .0 | .92 | -.1 | 1.13 | .54 | .36 | 5 v2_2 |
| 73 | 27 | 2.70 | 2.74 | -.36 | .42 | 1.12 | .4 | 1.11 | .4 | .84 | .09 | .36 | 6 v2_3 |
| 72 | 27 | 2.67 | 2.70 | -.19 | .40 | 1.20 | .7 | .88 | -.2 | 1.06 | .72 | .37 | 8 v3_2 |
| 71 | 27 | 2.63 | 2.66 | -.04 | .39 | 1.16 | .6 | 1.21 | .7 | .84 | .44 | .38 | 10 v3_4 |
| 69 | 27 | 2.56 | 2.59 | .25 | .37 | 1.40 | 1.4 | 1.28 | 1.0 | .51 | .02 | .40 | 14 v3_8 |
| 66 | 27 | 2.44 | 2.47 | .63 | .35 | 1.12 | .5 | 1.14 | .6 | .79 | .32 | .41 | 9 v3_3 |
| 61 | 27 | 2.26 | 2.27 | 1.19 | .33 | 1.03 | .2 | 1.03 | .2 | .92 | .36 | .43 | 2 v1_2 |
| 58 | 27 | 2.15 | 2.15 | 1.50 | .32 | .59 | -1.9 | .61 | -1.8 | 1.64 | .64 | .44 | 7 v3_1 |
| 57 | 27 | 2.11 | 2.12 | 1.60 | .32 | .74 | -1.1 | .75 | -1.1 | 1.36 | .41 | .44 | 11 v3_5 |
| 69.7 | 27.0 | 2.58 | 2.61 | .00 | .41 | 1.02 | .1 | .96 | -.1 | | .35 | | Mean (Count: 14) |
| 6.5 | .0 | .24 | .25 | .91 | .08 | .22 | .9 | .24 | .8 | | .23 | | S.D. (Population) |
| 6.7 | .0 | .25 | .26 | .94 | .08 | .23 | .9 | .25 | .9 | | .24 | | S.D. (Sample) |

Model, Populn: RMSE .42 Adj (True) S.D. .80 Separation 1.92 Strata 2.90 Reliability .79
Model, Sample: RMSE .42 Adj (True) S.D. .84 Separation 2.01 Strata 3.02 Reliability .80
Model, Fixed (all same) chi-squared: 73.4 d.f.: 13 significance (probability): .00
Model, Random (normal) chi-squared: 11.1 d.f.: 12 significance (probability): .52

According to Table 6, both Infit and Outfit Mean Square (Mnsq) values fall within the acceptable range of 0.5 to 1.5 logits, demonstrating the validity of the evaluation criteria.

Regarding item difficulty, the easiest area for advanced Iranian learners of Korean was found to be the "quantity of writing (V3_6)", with a difficulty estimate of -1.59. Conversely, the most challenging area was "accuracy of spacing and spelling (V3_5)", with a difficulty estimate of 1.60. This was followed by "appropriateness of vocabulary (V3_1)" with a difficulty estimate of 1.50 and "diversity of content (V1_2)" with an estimate of 1.19. The disparity between the easiest and most difficult areas is distinctly notable (see Figure 3).

The following are examples of the analysis conducted by raters on the answer sheets and descriptive content recorded in Excel files for the areas that advanced Iranian learners of Korean found most challenging: "accuracy of spacing and spelling (V3_5)", "appropriateness of vocabulary (V3_1)", and "diversity of content (V1_2)."

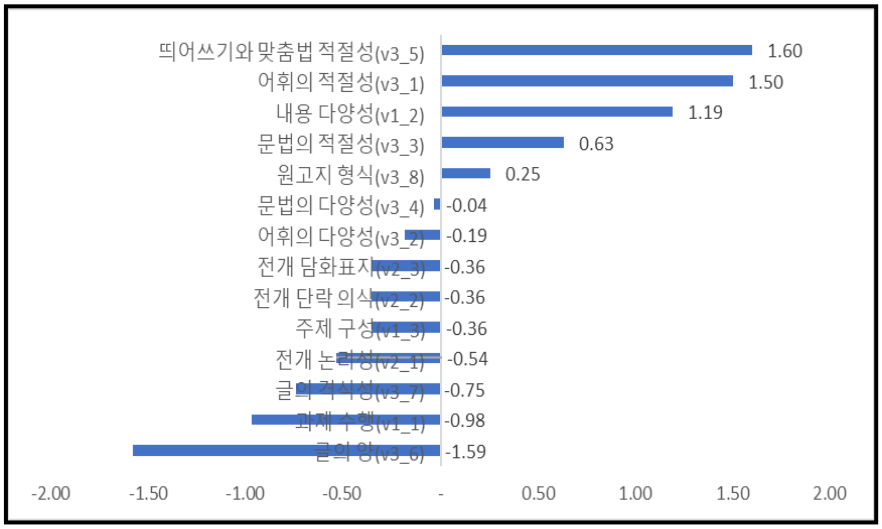



*Figure 3.* Results of the difficulty analysis of the evaluation scale.

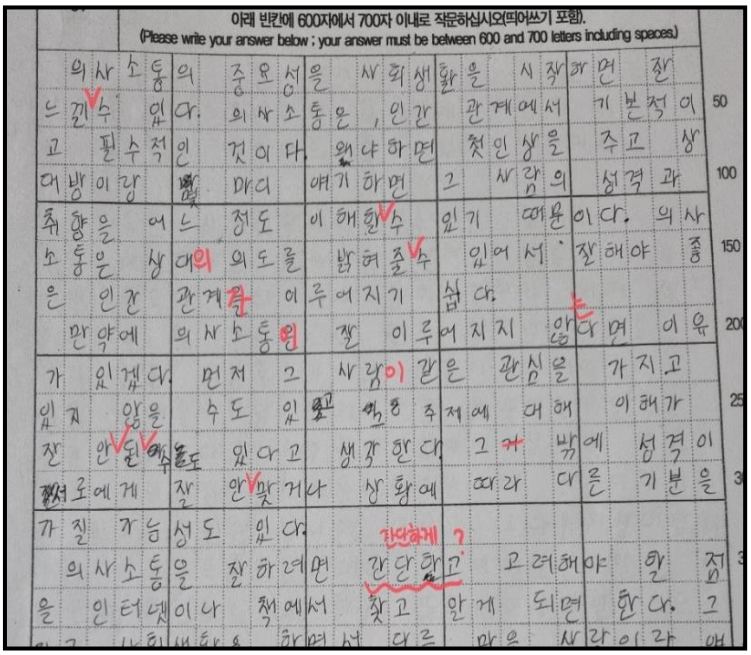



*Figure 4.* Example of "accuracy of spacing and spelling."

The most frequent errors were related to spacing. There were also frequent errors in the use of verb endings. Overall, the usage of spelling was inappropriate or did not conform to standard orthographic conventions. It appears that education on the proper use of spacing and spelling is necessary, starting from the basic stages of writing instruction (see Figures 3 & 4).

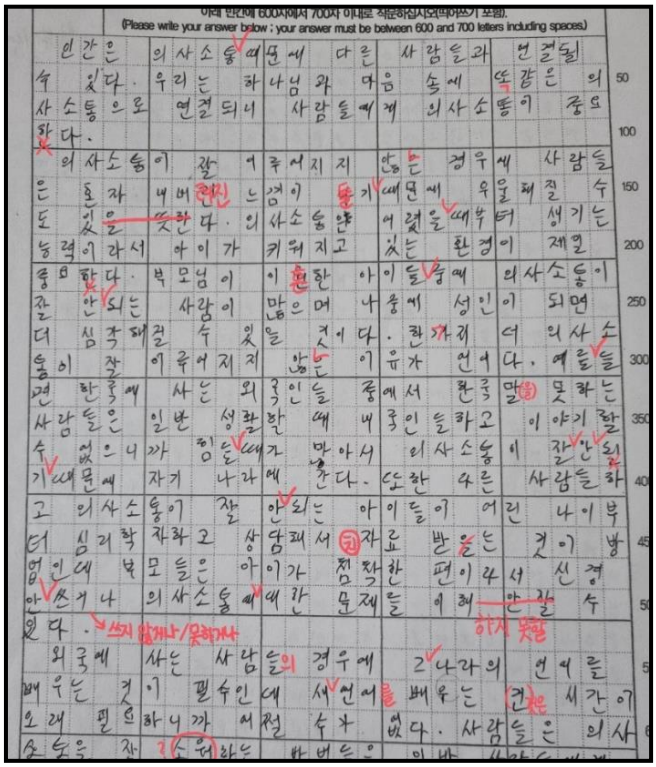


*Figure 5.* Example of "appropriateness of vocabulary".

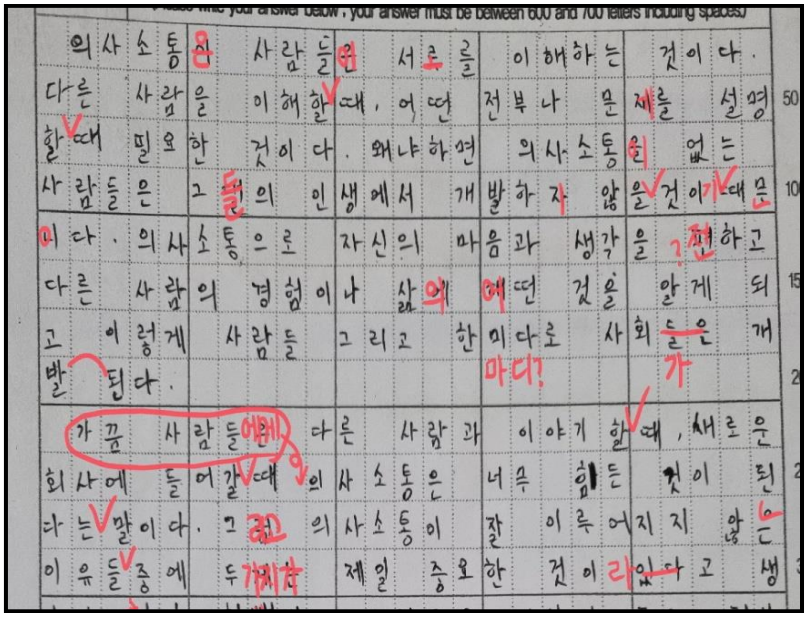


*Figure 6.* Example of "diversity of content".

Overall, there are difficulties in selecting vocabulary that is appropriate to the context, resulting in incomplete Korean sentences (see Figures 5 & 6). Additionally, although the writing may appear well-constructed at first glance, a detailed analysis reveals a tendency to choose simpler, beginner-level vocabulary rather than more appropriate Sino-Korean words due to the challenges in selecting suitable vocabulary.

Although the presented topic is addressed, the content could be more varied. While it may appear that a variety of content is being presented, the focus seems to be on writing a large quantity by repeating similar ideas, indicating a limited understanding of the topic. As a result, the development of the task content needs to be more cohesive, often appearing as a simple listing of points or displaying a disorganized structure.

## Conclusion and Discussion

The primary objective of this study was to identify the writing patterns of Iranian intermediate Korean learners in the TOPIK essay writing assessment to enhance their writing skills. Initially, the analysis revealed that Korean raters applied the TOPIK writing assessment criteria stringently. Using the FACETS program, the study evaluated the suitability and reliability of the writing assessment areas, demonstrating high validity with internal and external fit indices within the standard range of 0.5 to 1.2 logits.

The analysis indicated that the most challenging aspects of writing for Iranian intermediate Korean learners were listed as follows:

"Accuracy of spacing and spelling (V3_5)." This finding aligns with previous studies on writing errors in Korean essay content for international students, which reported high frequencies of particle and verb ending errors (Seo, 2014). Similarly, a study on spacing errors among Chinese Korean learners indicated that the spacing rules are particularly difficult (Kim, 2019). According to Kim's (2019) analysis, this is due to the difficulty of regularizing and generalizing space in Korean orthography. In contrast, Iranian spacing is routinely implemented according to well-defined norms. As a result, the Korean writing style, which recognizes tolerance for spacing, was the aspect that Iranian language learners found most problematic.

"Appropriateness of vocabulary (V3_1)." This is consistent with Seo (2014), who highlighted inappropriate vocabulary use and confusion with similar meanings as common errors in international students' Korean writing. Nguyen (2015) also noted that the diversity of vocabulary and expressions in Korean makes it difficult for foreign learners to use advanced-level vocabulary accurately, which is relevant to this study's findings.

"Diversity of content (V1_2)." This area was identified as challenging. Consistent with Nguyen (2015), the study found that limited and inappropriate understanding of topics hinders proper writing among Vietnamese Korean learners. This aligns with the qualitative evaluation of task analysis by the rest of this study, indicating a need for better guidance in accurately understanding topics to enhance content diversity.

Additionally, the task analysis by raters revealed issues with cohesion in developing task content. This phenomenon is similarly reported by (Shin, 2017), who noted that international students struggle with creating cohesive texts from subtext structures in academic writing. This study suggests that repetitive content with similar sentences and a lack of proper usage of intermediate or advanced vocabulary contribute to these issues. The raters observed that Iranian intermediate learners rarely use Sino-Korean vocabulary. Since Persian, an Indo-European language, follows modern Persian orthography, Iranian learners have limited background knowledge of Sino-Korean vocabulary (Shin, 2016). To enhance the effectiveness of Korean language education, it is fundamentally necessary to teach Hanja, which forms the basis of advanced vocabulary.

The discussion on the challenges of "accuracy of spacing and spelling," "appropriateness of vocabulary," and "diversity of content" suggests three main reasons for the difficulty in structuring and developing writing.

The Persian language curriculum for Iranian learners focuses on composition, dictation, reading, and grammar, providing limited opportunities for practicing argumentative writing.

Language education's emphasis on learning vocabulary, reading, listening, and speaking leads to relatively low investment in writing practice and feedback.

These factors likely contribute to the relatively low scores in argumentative writing among Iranian intermediate learners in the TOPIK writing test.

## Policy Implications

The findings of this study underscore the critical need for targeted educational policies and strategies to enhance the writing proficiency of Iranian intermediate Korean learners. Given the identified challenges in areas such as spacing and spelling accuracy, vocabulary appropriateness, and content diversity, policymakers, educators, and curriculum developers must implement comprehensive measures to address these issues. The following policy recommendations are proposed to improve the effectiveness of Korean language instruction and support the academic success of Iranian learners:

1. Curriculum development: Integrate comprehensive writing courses into the primary and secondary education curriculum, emphasizing argumentative writing skills to provide learners with early and consistent training. This integration will help learners build a strong foundation in writing from an early stage.

2. Teacher training: Enhance professional development for Korean language instructors to equip them with effective strategies for teaching spacing, spelling, and advanced vocabulary, particularly to learners with different linguistic backgrounds. Systematic training related to spacing and spelling accuracy is essential (Ji, 2018).

3. Resource allocation: Increase investment in web-based tools and resources that offer real-time feedback on writing tasks, enhancing learners' ability to self-correct and improve their writing skills. Utilize web-based programs that provide immediate feedback on spelling and grammar checks in the teaching-learning process.

4. Cultural and linguistic integration: Develop educational materials that bridge the gap between Korean and learners' native languages, such as incorporating Hanja education for learners from non-Sino-Korean linguistic backgrounds. Additionally, training programs for argumentative writing targeted at foreign learners are necessary to enhance vocabulary appropriateness and content diversity.

5. Continued education and training: Encourage continued education to establish basic writing structures and provide training in using Korean genre and rhetorical patterns in argumentative writing. This ongoing education will ensure that learners develop the necessary skills to construct cohesive and compelling arguments in their writing.

By implementing these strategies, educational institutions can significantly improve the writing abilities of advanced Iranian learners, thereby enhancing their overall proficiency in the Korean language and supporting their academic and professional success.

## Limitations and Future Studies

This study has several limitations. Firstly, due to the small sample size, the generalizability of the writing scores to a broader population is restricted. Secondly, this study focused on part 52 of the writing section in the 54th TOPIK, limiting the extent to which these findings can be generalized to scores from other rounds of the

test. Consequently, future research should include a more comprehensive analysis of the content, format, scoring process, and volume of the texts to enhance the robustness of the findings.

## References


Ahn, J. (2012). A study on analyzing learner text structures for academic Korean writing education. *Journal of Writing Research, 15*, 101-126.

Batjargal, E. (2017). *Analysis of argumentative writing texts and educational strategies for mongolian learners of Korean for academic purposes*. Ewha Womans University Graduate School of International Studies.

Choi, E. J. (2009). A study on the Korean wave in Iran: Focusing on the broadcasting of "Dae Jang Geum" (Master's thesis, Kyung Hee University).

Choi, S. K. (2016). A study on the evaluation patterns of middle school Korean language teachers' book review assignments using the multidimensional Rasch model. *Journal of Learner-Centered Curriculum and Instruction, 16*(1), 515-538.

Dashdorj, S., & Sodnomdorj, E. (2020). Writing performance patterns of Mongolian learners in the writing section of the Test of Proficiency in Korean (TOPIK). *Korean Language Education, 31*(3), 339-363.

Derakhshan, A., & Karimian, S. R. (2020). An investigation of the Iranian EFL learners' perceptions towards the most common writing problems. *Sage Open, 10*(2), 2158244020919523.

Engelhard, Jr, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences*. Routledge.

Jeong, M. K. (2015). Analysis of the mediating effects of career decision self-efficacy and career preparation behavior in the relationship between career barriers and career decision among college students. *Korean Journal of Education, 21*(3), 85-106.

Ji, M. G. (2018). The structural relationship between Korean writing and Hangul spelling for Mongolian Korean learners. *Linguistics, 26*(4), 225-243.

Khodabandeh, F., Jafarigohar, M., Soleimani, H., & Hemmati, F. (2014). Investigating Iranian EFL writing problems and examining back transfer. *Asian EFL Journal, 16*(1), 227-270.

Kim, B. M. (2009). A study on the grammatical expressions in persuasive writing of middle school students (Master's thesis, Korea University).

Kim, C., Lee, J. H., Pyun, D. O., & Byon, A. S. (2022). KSL assessment: The TOPIK, achievement tests, and research trends. In *The Routledge Handbook of Korean as a Second Language* (pp. 493-517). Routledge.

Kim, J. K. (2005). A study on methods for improving essay writing skills (Master's thesis, Ewha Womans University Graduate School).

Kim, J. S. (2000). A preliminary study for designing Korean language curriculum for academic purposes: Focusing on curriculum for university students. *Korean Language Education, 11*(2), 1-19.

Kim, M. Y. (2004). A study on improving the reliability of essay evaluation: Focusing on Evaluator Meetings (Master's thesis, Korea University).

Kim, S. S. (2015). *Theories and practices of Korean writing education*. Gyeongjin Publishing.

Kim, Y. I. (2019). A study on spacing errors in the writing of intermediate Chinese learners. *Uirimalgeul,* 82, 237-274.

Knoch, U. (2009). *Diagnostic writing assessment: The development and validation of a rating scale* (Vol. 17). Peter Lang.

Lee, D. H. (2003). A study on curriculum design for academic purposes in Korean language education through needs analysis (Unpublished master's thesis, Yonsei University).

Lee, H. (2012). An analysis of scoring reliability based on pronunciation category assessment methods in Korean speaking performance evaluation: Using the multifaceted Rasch model. *Korean Language Education as a Foreign Language, 37*, 325-355.

Lee, M. H. (2017). Error patterns in the TOPIK (Test of Proficiency in Korean) writing of Korean language learners: Focus on Task 53 (Descriptive writing). *Korean Language and Culture, 14*(2), 231-254.

Lee, M. H. (2010). Content system of genre-based Korean writing education. *Foreign Languages Education, 17*(3), 463-485.

Linacre, J. (1998). *FACETS 3.17 Computer program.* Chicago, IL: MESA Press.

Linacre, J. M. (1996). *A user's guide to facets: Rasch measurement computer program (version 3.0)*. Chicago, IL: MESA Press.

Lumley, T., & McNamara, T. F. (1995). Rater characteristics and rater bias: Implications for training. *Language Testing, 12*(1), 54-71.

Ma, G. H. (2019). A study on error patterns in answer writing for the Korean Language Proficiency Test (TOPIK) No. 54 (Master's thesis, Chung-Ang University Graduate School).

Marashi, S. M. (2020). Impact of teaching cohesive devices to Iranian EFL learners: Case of IELTS writing task 2 examination. *Online Submission, 6*(2), 115-126.

McNamara, T. F. (1996). *Measuring second language performance*. Longman.

Min, B. G. (2001). The current status of argumentation theory and tasks in Korean language education. *Korean Language Education Research*, 12.

Min, B. G. (2004a). A study on the content of argumentation education: Based on analysis of writing and discussions of 6th, 8th, and 10th grade learners (Doctoral dissertation, Seoul National University).

Min, B. G. (2004b). An analysis of the use of argumentation schemas in the production process of argumentative texts and its educational implications. *Korean Language Education Research,* 18.

Nam, J. H. (2011). A study on developing textbooks for academic purposes Korean essay writing (Master's thesis, The Catholic University of Korea Graduate School).

Nasrollahi, V. R. (2020). A study on pronunciation instruction of Korean diphthongs for Iranian learners (Master's thesis, Seoul National University Graduate School).

Nguyen, T. H. S. (2015). A study on Korean writing education for Vietnamese learners: Focusing on argumentative writing (Doctoral dissertation, Seoul National University).

Nimehchisalem, V., Abbasi, M. M., Ebrahimzadeh, A., & Kalajahi, S. A. R. (2015). Iranian English as a foreign language (EFL) learners' argumentative writing performance in private language institutes. *Asian Social Science, 11*(15), 96.

Oh, J. H. (2015). Issues and improvement measures in the "writing" assessment of the Korean Language Proficiency Test (TOPIK). *Korean Language and Literature Studies,* 32, 77-98.

Park, H. D. (2009). A study on essay evaluation and revision instruction methods (Doctoral dissertation, Yeungnam University Graduate School).

Park, Y. H. (2016). A study on the writing patterns of Korean language learners by proficiency level through analysis of TOPIK writing test results (Master's thesis, Inha University Graduate School).

Salmani Nodoushan, M. A. (2018). Toward a taxonomy of errors in Iranian EFL learners' basic-level writing. *Online Submission, 12*(1), 61-78.

Seo, S. B. (2014). Analysis of Korean writing errors by international students: A case study of university students' essay writing in a Korean composition contest. *Uri Malgeul, 62*, 127-157.

Seo, S. H. (2003). A study on the criteria for setting writing assessment standards (Master's thesis, Korea University).

Seo, S. H. (2008). A study on setting criteria for writing assessment through factor analysis (Doctoral dissertation, Korea University).

Shin, D., & Seol, H. (2005). Analyzing scoring data using NEW FACETS. *Foreign Languages Education, 12*(2), 191-212.

Shin, Y. (2016). *Persian grammar*. Myeongji Publishing.

Shin, J. (2017). A study on academic text writing education for international students. *Text Linguistics,* 43, 117-136.

Song, W. (2020). Analysis of writing performance patterns of Chinese Korean language learners using the Rasch model (Doctoral dissertation, Korea National University of Education).

The Ministry of Education. (2023). *Statistics on international students in domestic Higher Education Institutions*. Retrieved from https://www.moe.go.kr/main.do?s=moe

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Yan, W., & Cheng, L. (2015). How language proficiency contributes to Chinese students' academic success in Korean universities. *Language Testing in Asia,* 5, 1-19.

Yang, I. Y. (2008). A study on the structural analysis of argumentative texts in elementary school textbooks and their educational application (Master's thesis, Kyongin National University of Education).

**Appendix**

| **1. Evaluation Category: Content and Task Performance** | | | |
|---|---|---|---|
| **Evaluation Content** | **Evaluation Items** | **Scoring Criteria** | **Score** |
| (1-1) **Task Performance**: Has the given task been thoroughly completed? | Why is communication important? What are the reasons for poor communication? What are the methods to facilitate effective communication? | If three questions are answered. | 3 |
| | | If two questions are answered. | 2 |
| | | If one question is answered. | 1 |
| (1-2) **Diversity of Content**: Has the given content been expressed diversely and richly? | Has the importance of communication and the methods for effective communication been expressed richly and diversely? | If the importance of communication is well expressed and more than three methods are described. | 3 |
| | | If the importance of communication is well expressed and two methods are described. | 2 |
| | | If the importance of communication is poorly expressed and one method is described. | 1 |
| (1-3) **Topic Composition**: Is the content composed around the topic? | Is the content composed related to the topic of 'the importance of communication and its methods'? | The theme is clear and consistently articulated. | 3 |
| | | There are one or two irrelevant or unnecessary parts. | 2 |
| | | There are three or more irrelevant or unnecessary parts, or the content only partially addresses the topic. | 1 |
| **2. Evaluation Category: Structure of the Writing** | | | |
| (2-1) **Logical Development**: Is the composition clear and logical? | Was it written logically? | Written logically without any errors. | 3 |
| | | Contains 1-2 logical errors. | 2 |
| | | Contains 3-4 logical errors. | 1 |
| (2-2) **Paragraph Structure**: Is the paragraph structure appropriately organized according to the content of the writing? | Is there a clear paragraph structure with an introduction, body, and conclusion? | Written with clear paragraphs for introduction, body, and conclusion. | 3 |
| | | Satisfies two out of the three (introduction, body, conclusion). | 2 |
| | | Attempted to use the introduction, body, and conclusion structure but wrote it as a single paragraph. | 1 |
| (2-3) **Discourse Markers**: Were discourse markers appropriately used to aid logical development and organize the content systematically? | Were discourse markers appropriately used to aid logical development and organize the content systematically? | Used appropriate cohesive devices between paragraphs (e.g., first, next, finally) and smoothly connected sentences with conjunctive adverbs. | 3 |
| | | Did not use cohesive devices between paragraphs, but the logical flow is not hindered. | 2 |
| | | Did not use cohesive devices between paragraphs, and the logical flow was not hindered. | 1 |
| | | Did not use cohesive devices between paragraphs, and the logical flow is hindered. | 0 |
| **3. Evaluation Category: Language Use** | | | |
| (3-1) **Appropriateness of Vocabulary** | Did the writer choose appropriate grammar and vocabulary and use them diversely and richly? | There are almost no vocabulary errors (1-3 errors). | 3 |
| | | There are occasional vocabulary errors (4-6 errors). | 2 |
| | | There are generally vocabulary errors (7-9 errors). | 1 |
| | | There are significantly many vocabulary errors (10 or more). | 0 |
| (3-2) **Variety of Vocabulary** | Were a variety of vocabulary words used? | Used a variety of intermediate or advanced vocabulary and expressions very well (10 or more). | 3 |
| | | Used a variety of intermediate or advanced vocabulary and expressions (7-9). | 2 |
| | | Did not use a sufficient variety of intermediate or advanced vocabulary and expressions (4-6). | 1 |
| | | Did not use a variety of intermediate or advanced vocabulary and expressions (3). | 0 |
| (3-3) **Appropriateness of Grammar** | Was appropriate grammar used? | There are not many grammar errors (1-3 errors). | 3 |
| | | There are occasional grammar errors (4-6 errors). | 2 |
| | | There are generally grammar errors (7-9 errors). | 1 |
| | | There are significantly many grammar errors (10 or more). | 0 |

| | | | |
|---|---|---|---|
| (3-4) **Variety of Grammar** | Was a variety of grammatical structures used? | Used a variety of intermediate or advanced grammatical structures very well (5 or more). | 3 |
| | | Used a variety of intermediate or advanced grammatical structures (3-4). | 2 |
| | | Did not use a sufficient variety of intermediate or advanced grammatical structures (1-2). | 1 |
| | | Did not use intermediate or advanced grammatical structures. | 0 |
| (3-5) **Appropriateness of Spacing and Spelling** | Is the use of spacing and spelling accurate? | There are almost no spacing and spelling errors. | 3 |
| | | There are occasional spacing and spelling errors, but they do not hinder the communication of meaning. | 2 |
| | | There are generally spacing and spelling errors, but they hardly hinder the communication of meaning. | 1 |
| | | Spacing and spelling are poor, and the content includes irrelevant information in a 600–700-word text. | 0 |
| (3-6) **Quantity of Writing** | Was the quantity of writing appropriate? | Excluding copying the prompt or irrelevant content, wrote 600-700 words. | 3 |
| | | Excluding copying the prompt or irrelevant content, wrote 500-599 words. | 2 |
| | | Excluding copying the prompt or irrelevant content, wrote 400-499 words. | 1 |
| (3-7) **Formality of Writing** | Was the formal language used instead of colloquial language? | Formal language was used well. | 3 |
| | | There are occasional uses of colloquial language (1-2 instances). | 2 |
| | | There are generally uses of colloquial language (3-4 instances). | 1 |
| | | There are many uses of colloquial language (5 or more instances). | 0 |
| (3-8) **Manuscript Format** | Was the manuscript format well considered? | The manuscript format was well adhered to. | 3 |
| | | There are occasional errors in the manuscript format (1-2 instances). | 2 |
| | | There are generally errors in the manuscript format (3-4 instances). | 1 |
| | | There are significantly many errors in the manuscript format (5 or more instances). | 0 |